

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:01:57 ; Search time 35.72 Seconds

(without alignments)
3521.177 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013
Sequence: 1 PGLTIAIEQALAANSALA.....FTETIKPVLDTNSIIQRIE 1698

Scoring table: BLOSUM62

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%

Listing first 45 summaries

Database : A_Geneseq_1101:*

1:	/SID58_gcgdata1/geneseq/geneseqp/AA1980.DAT *
2:	/SID58_gcgdata1/geneseq/geneseqp/AA1981.DAT *
3:	/SID58_gcgdata1/geneseq/geneseqp/AA1982.DAT *
4:	/SID58_gcgdata1/geneseq/geneseqp/AA1983.DAT *
5:	/SID58_gcgdata1/geneseq/geneseqp/AA1984.DAT *
6:	/SID58_gcgdata1/geneseq/geneseqp/AA1985.DAT *
7:	/SID58_gcgdata1/geneseq/geneseqp/AA1986.DAT *
8:	/SID58_gcgdata1/geneseq/geneseqp/AA1987.DAT *
9:	/SID58_gcgdata1/geneseq/geneseqp/AA1988.DAT *
10:	/SID58_gcgdata1/geneseq/geneseqp/AA1989.DAT *
11:	/SID58_gcgdata1/geneseq/geneseqp/AA1990.DAT *
12:	/SID58_gcgdata1/geneseq/geneseqp/AA1991.DAT *
13:	/SID58_gcgdata1/geneseq/geneseqp/AA1992.DAT *
14:	/SID58_gcgdata1/geneseq/geneseqp/AA1993.DAT *
15:	/SID58_gcgdata1/geneseq/geneseqp/AA1994.DAT *
16:	/SID58_gcgdata1/geneseq/geneseqp/AA1995.DAT *
17:	/SID58_gcgdata1/geneseq/geneseqp/AA1996.DAT *
18:	/SID58_gcgdata1/geneseq/geneseqp/AA1997.DAT *
19:	/SID58_gcgdata1/geneseq/geneseqp/AA1998.DAT *
20:	/SID58_gcgdata1/geneseq/geneseqp/AA1999.DAT *
21:	/SID58_gcgdata1/geneseq/geneseqp/AA2000.DAT *
22:	/SID58_gcgdata1/geneseq/geneseqp/AA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	%	Length	DB	ID	Description
1		9001	99.9	1598	20	AAV31381	HEV-OS1 ORF1 prote	
2		8812.5	97.8	1708	20	AAW93408	Swine HEV ORF 1 pr	
3		8753.5	97.1	1708	20	AAV93404	HEV-OS2 ORF1 prote	
4		7456	82.7	1693	17	AAW91813	Hepatitis E virus	
5		7448	82.6	1693	17	AAW80196	Protein encoded by	
6		7448	82.6	1693	21	AAW824119	Hepatitis E virus	
7		7448	82.6	1693	22	AAW65522	HEV-Burma strain v	
8		7439	82.5	1693	19	AAW81519	Hepatitis E virus	
9		7437	82.5	1693	19	AAW76368	Hepatitis E virus	
10		7437	82.5	1693	19	AAW71209	Protein encoded by	
11		7436	82.5	1693	15	AAW51264	HEV strain protein	

12	7400	82.1	1.693	12	AAR44618
13	2057	22.8	4.31	12	AAR44615
14	2057	22.8	4.31	19	AAB80195
15	2057	22.8	4.31	19	AAM71208
16	2057	22.8	4.31	21	AAB24118
17	2057	22.8	4.31	22	AAB82521
18	709	7.9	1.38	20	AAB31406
19	705	7.8	1.38	20	AAB31408
20	701	7.8	1.38	20	AAB31395
21	504	5.6	1.704	15	AAR96657
22	439	4.4	1.615	19	AAM39276
23	393	4.5	1.65	22	AAB31997
24	393.5	4.4	1.64	22	AAB32000
25	374.5	4.2	2.205	16	AAR99048
26	365	4.0	78	14	AAR44727
27	360	4.0	1.233	19	AAW19335
28	350	4.0	1.233	19	AAW45536
29	268.5	3.0	1.14	22	AAB31995
30	259	2.9	1.14	22	AAB31998
31	213.5	2.4	2161	20	AAB87724
32	193	2.1	1116	15	AAR06088
33	185	2.1	2431	13	AAR25138
34	184	2.0	1390	18	AAW21636
35	184	2.0	2237	21	AAW58148
36	180	2.0	501	22	AAB81999
37	178	2.0	34	14	AAR44728
38	178	2.0	501	22	AAB31996
39	177.5	2.0	525	21	AAW1964
40	176.5	2.0	935	21	AAW07570
41	165.5	1.8	265	20	AAW07239
42	163	1.8	1066	21	AAW07561
43	162	1.8	1456	14	AAR54129
44	161.5	1.8	1128	11	AAR05107
45	161	1.8	3266	21	AAW2491

ALIGNMENTS

Accession	Protein	Location/Qualifiers
AAV31381	AAV31381 standard; Protein; 1698 AA.	
XX	AAV31381;	
AC	12-OCT-1999 (first entry)	
XX	HEV-US1 ORF1 protein.	
DE		
XX	Hepatitis E virus; HEV; binding partner	
KW	vaccine; passive immunisation.	
XX	Hepatitis E virus.	
OS		
XX		
FH	Key	
FT	Misc-difference	174
FT		/label= unknown
FT		/note= "encoded by ATR
FT		
FT	Misc-difference	363
FT		/label= unknown
FT		/note= "encoded by RTG"
FT		
FT	Misc-difference	1088
FT		/label= unknown
FT		/note= "encoded by YTC"
FT		
FT	Misc-difference	1131
FT		/label= unknown
FT		/note= "encoded by YTC"
FT		
FT	Misc-difference	1217
FT		/label= unknown
FT		/note= "encoded by CRC
FT		
FT	Misc-difference	1389
FT		/label= unknown
FT		/note= "encoded by GMC"

HEV isolate 2015-1	Protein encoded by HEV ORF 1 open reading frame
HEV isolate 1440013	Hepatitis E virus protein encoded by HEV ORF 1 open reading frame
HEV peptide z12-orn	Sequence of Helicobacter pylori
Rubella virus RA27/3	Amino acid sequence of Rubella virus
Amino acid sequence infectious rubella	Infectious rubella virus amino acid sequence
HEV ORF1 peptide (Nidaurelia beta-VH)	HEV ORF1 peptide (Nidaurelia beta-VH)
Nidaurelia beta-VH	Amino acid sequence of Nidaurelia beta-VH
Amino acid sequence	Amino acid sequence of Rubella virus
Rubepritis stem pili	Rubepritis stem pili
SFV4 non-structural	SFV4 non-structural protein
Graevia leafroll	Graevia leafroll virus
GRAV-3 polypeptid	GRAV-3 polypeptide
Amino acid sequence	Amino acid sequence of Rubella virus
HEV ORF1 peptide (Amino acid sequence	HEV ORF1 peptide (Amino acid sequence of Rubella virus)
Pknox-1 RNA-dependent	Pknox-1 RNA-dependent protein
Rubepritis stem pili	Rubepritis stem pili
Protein encoded by Pox replicase	Protein encoded by Pox replicase
Human OREF ORF2255	Sequence encoded by Human OREF ORF2255

XX WO9919732-A1.
PN
PD 22-APR-1999.
XX
PF 15-OCT-1998; 98WO-US21941.
XX
PR 15-OCT-1997; 97US-0061199.
XX
PA (ABBO) ABBOTT LAB.
PI Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;
DR MPI, 1999-288017/24.
XX N-PSDB: AAZ00195.
PT Detection of United States Isolates of hepatitis E virus

XX
PS Claim 6; Page 175-180; 260pp; English.

CC The invention provides a method for detecting a US (sub)type hepatitis E
CC virus (US-HEV), or its naturally occurring variants in a sample by
CC treatment with a binding partner specific for a marker of the virus, and
CC then detecting any complex formed. The method is used to diagnose
CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
CC are useful in vaccines or for passive immunisation. The polypeptides are
CC also used to raise specific antibodies (useful as immunoassay reagents).
CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
CC usual hybridisation and amplification assays for detecting infection. The
CC present sequence represents a HEV-US1 ORP1 protein.

XQ Sequence 1698 AA;

Query Match	99.9%	Score 9001;	DB 20;	Length 1698;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1698; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	PGITTAIEOALAAANSALANAVVVRFLSVQTEIILINMQPOLVFRREVLMNHPIQR	60
Db	1	pgittalegaalaanaansalanavvrpflsvrqceiilnmqprqlvfrrevlmnhpiqr	60
-QY	61	VIHNEEOYCARAGRCLEVCAMHRRSINDPNVLHRCFLRPVGDVORWTSAPPRGAAN	120
Db	61	vihmelegycaragrclevghabhrsindpnvlhrcflrpvgdvorwtsaptrgpaan	120
QY	121	CRRSALRCLPADRFYPCDGFSCAPAAEENGVALTSLHDMPADVAAMAHGCTRILYAA	180
Db	121	crrsalrlgpradrtfctcdfsrfcafaeengvalyslhdmpadvaamearhgxtrilyaa	180
QY	181	LHLPREVLPRGYHTTYSYLIIHDGDRAVVTYEGDTSAGINHDSILIRAMIRTKTVGDH	240
Db	181	lhlppevlprpgyhttsyllihhgdravvtyegdtsgaynhdssilrawlrctkivgdh	240
QY	241	PLVIERVVAICHHVLLLTAAPEESPMPYVYPRSTEVYRSIGPGSPSLPFSASCTK	3000
Db	241	plviervvaicshvlllttaapeespmpyvprrstevyrslgpgspslfpfsactk	3000
QY	301	STFHAVPVHIMDRMLEGATLDDOAFCCSRIMTYLRGISRYTVGALVANEGMASSEDAL	3606
Db	301	stfhavpvhidrmlefgatlddofccsrilmtylrgisryktvgalvanegmasedal	3606
QY	361	TAXITAAVITTCHORLYLRTQAIISGMRMRLGYENHAKRTILRYSMLFPKSGRODTPGQLO	4200
Db	361	taxitaaylvttchorylrtlqaiisgmrrlgyehaqrktilrlyslfeksgrdytpgqrlq	4200
QY	421	FYAOCRMLSGAFHIDBRVLVTFDESVCRCRTFLKVAAGFCCPMRMJAGCECFLPEAE	4800
Db	421	fyagcrrmlsghfidbrvlvtfdesvrcrctflkvaagfcccmmrjagceccflpeae	4800
QY	481	GLVGDHGDNEAYGCGSEYDPAEPANHDVSGTYAVHGOLEALRYALNVPDIDAAARSLT	5400

Dh	481	qlyvdghndneaygsevdpraeahldvsgtyavhghnqlealyrzalnypgdlaarastlt	540
Qy	541	atvelvaspdrlectrvlgnkfrfrrtyvdcahlleangpeqoyvlspdasrqskgagslsht	600
Dh	541	atvelvaspdrlectrvlgnkfrfrrtyvdcahlleangpeqoyvlspdasrqsngashtlt	600
Qy	601	yelrpaglqvnsngldcmnfprpggapsaaacgfuaacsalryvnrprrtohslitgltm	660
Dh	601	yelrpaglqvnsngldcmnfprpggapsaaacgfuaacsalryvnrprrtohslitgltw	660
Qy	661	lhpegllgiefppspghimesanpfcgegtlylrwtstgsfssdpspppnaaPAMAAPG	720
Dh	661	lhpegllgiefpfsppghlimesanpfcgegtlylrwtstgsfssdpspppnaaPAMAATG	720
Qy	721	lphstppvsdiwvlrppseefoydAAPVPAPDPAGlpcgvvyltpppprvhkslppps	780
Dh	721	lphstppvsdiwvlrppseefoydAAPVPAPDPAGlpgvrvltpppprvhkslppps	780
Qy	781	rnrbrlvtprdgaaVVVvGSLFESDCOMLVNASSpGRPCGGCLHAFYORFPAVPTEFI	840
Dh	781	rnrlllylrypdgaVVvYvagslfesdcwlvnaaspgrprp9glchafyqrffpaatpcefl	840
Qy	841	MREGLIAVTLPRPFIHAHVAPDVRVEONPKRLAEARFECSSRRGGAAPPLSGIYQVPV	900
Dh	841	mreglaaytltprrpllnhavapdyveqnpkrlaeayreccsrgrgcaaplllgsjlyqvpv	900
Qy	901	SLUSDAMERNHRpGDELyLTERPAA NFPAKPAQVYLTTEDTARTANLALETDAATEVG	960
Dh	901	slsfdawernhrpddelylterpaanffeankpqavlltledartalanaleidaatevg	960
Qy	961	RACGCGITSGIYVHOFTACVPGSGSRSIIOGDDVVVVVPRREL RNSRRRGFAAPFH	1020
Dh	961	raccgcltspglvmyqtlacvpgpsgsrslsgqddvvnvvprrtelrnsrrrrgtlaatfph	1020
Qy	1021	TAAAVTTIGRRVVIDEAPSLPRHLHLHMO RASSVHLHGDPNIPAIDEHAGLVPAIRPE	1080
Dh	1021	taavtltgrrrvvideapslprhlhlhmqraassvhlhgpnqjpaidehaglvpatrpe	1080
Qy	1081	LAPTSMMXVTHRCPADYCELI RGAAYPKIOTTSVYLSLEFWNEPAIGOKUYLTQAANKANP	1140
Dh	1081	lapswwxvthrcpadycelllrgaypkltcslrvlsrlfwnepalqgklvxqcaakaanp	1140
Qy	1141	GATVHEAGOGTFEMTIIINTAARGLIOSSRHAIVALTRHEKVCVILDAAGLREVI	1200
Dh	1141	gatlvehaagacflfettcllatacadarglqgsrahaivaltrhkevcvllldagjlrrevgl	1200
Qy	1201	SDVIVNNFPLAGGEVGHXHSRVP RKNPNPDONLGTLOAFPPSCOISAYHQLEELGHRAP	1260
Dh	1201	sdivnnmfplaggevgxhrrsvprknppndonlgtloafppscoisayhqleelghrrap	1260
Qy	1261	VAAVLPPCPPELEGGLYMPDELTVSDSVLFEJLTDVHCMAAPSOKRAVUSTLVGRYGR	1320
Dh	1261	vaavlppcppeleggllympdeltvssdvsvlfejltdvhcmaapsqkxavustlvgryr	1320
Qy	1321	RTKTYEAASHDVRSIARFPTIGPVOATTCCELYELVEAMVKGODGSVYLELDCNNDV	1380
Dh	1321	rtktyeaashdvrsiarfptigpvoatccelyelveamvkgodgsavleldcncdv	1380
Qy	1381	SRIFFPKXCKKFTTETIANGKVGGCISAMSTFCALGPMFPAIIEKITALLPRNIFY	1440
Dh	1381	sriffpkxckkfttgetiahgkvvgglsawsktfcaifgprfratekeiallpprnify	1440
Qy	1441	GDAAEESVFAAAVSGAGSCWPFENDSPEDSTONNESLGLCVCVMEECGMPOMLIRLYHL	1500
Dh	1441	gdayeesvfaaavsgagscnwvfendtseldstqgnfslglcvmeecgmppwllrlyhl	1500
Qy	1501	VRSAWILLQAPKESIKGFWMKHSGEPTLLMNTVYMMMAIIANCYBFRDRRVAAFKCDSDV	1560
Dh	1501	vrsawillqapkesikgfwmkhsgeptllmntvwmmailahcyefdrfrvaafkcdsdsv	1560
Qy	1561	LCSPYSRNRMAALLINGCGILKVDVRRPRTGLVAGVVVVAAGLCTRDVVRFGRGRISERKMG	1620
Dh	1561	lcsdyrsrnrmaalllagcgilkvdvrrprtyglvavvvvaaglyctrdvvrfgrgrisekmw	1620

```

OY 1621 PEPERAEQLRLAVCFPLKGLTINVAQCVDSRVYGVSPGLVHNLIGMLQTLADKAHFT 1680
Db 1621 pperaeqrlrlavcfdflrgltnvavcvdsvrvygvspglvhnligmqladgkahft 1680
OY 1681 ETIKRPVLDLTNSIIORVE 1698
Db 1681 etikrpvldltnsiqrve 1698

RESULT 2
ID AAM93408
AC AAM93408 standard; Protein; 1708 AA.
XX
XX AAM93408;
XX
XX 11-JUN-1999 (first entry)
XX
XX Swine HEV ORF 1 protein.
XX
XX Swine hepatitis E virus; HEV; cross-reaction; antibody; human; therapy;
XX vaccine; immunise; infection; detection; diagnosis; prevention.
XX
XX Hepatitis E virus.
XX
XX WO9904029-A2.
XX
XX 28-JAN-1999.
XX
XX 17-JUL-1998; 98WO-US14665.
XX
XX 18-JUL-1997; 97US-0053069.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Emerson SU, Meng X, Purcell RH;
XX
XX WPI: 1999-132270/11.
XX
XX New isolated swine hepatitis E virus - used to develop products for
XX the diagnosis, prevention and treatment of hepatitis E virus
XX infection in mammals, particularly humans
XX
XX Example 1; Fig 6D-J; 70pp; English.
XX
XX This invention describes a swine hepatitis E virus (HEV) and its natural
XX mutants which are capable of cross-reacting with antibodies reactive
XX with a human HEV strain or natural mutants. The HEV and the proteins
XX can be used in vaccines for immunising against HEV infection. The swine
XX HEV can be used in humans to prevent possible infection by human HEV. The
XX swine HEV can also be used as a therapeutic treatment for infection by
XX other strains of HEV. The swine HEV can also be used for the production
XX of antibodies which can be used in therapy, detection and diagnosis. The
XX products can also be used for determining the susceptibility of cells or
XX organs to infection with swine HEV. The swine HEV is particularly useful
XX for the development of agents for the prevention, treatment and detection
XX of human HEV because it is not a human virus and thus can be handled both
XX experimentally and clinically without fear of severe infection and/or
XX contamination.
XX
XX Sequence 1708 AA;

Query Match 97.8%; Score 8812.5; DB 20; Length 1708;
Best Local Similarity 97.8%; Pred. NO. 0;
Matches 1661; Conservative 7; Mismatches 30; Indels 1; Gaps 1;

OY 1 FGIITAIQDAALLAANSAALANAVVVRPEFSRQTEILINLKMOPROLYRFEVYLMNHPRIOR 60
Db 10 pgiitaiqdaaalaanaalanaavvrpfisrvtetlilnlmgpqlvfrfevylmnpriqr 69
OY 61 VIHNLEQYCARAGRCLEVGAAHPSINDNPVNLHRCFLRPYGRDQRMYSAPTRGPAA 120
Db 61 vihnleqycaragrclevgaaahpsindnpvnlhrcflrpygrdqrmysaptrgpaa 120

```

D	b	70	aimhel	eqyctaragccolevgahprfi	indprnvllhncflirpygrdqwy	sapltgrpaan	129					
Q	y	121	CRSALGKLEPADTTCYCFDFESRCAFAETGVALYSLHDI	IMPADVAEAMAR	RGXTRLYAA	180						
D	b	130	crisalrglprpvdrtlycfdfgfsrcafaetgvalylhndi	mpdvaeamar	ngmtrclya	189						
Q	y	181	LHLPPEVLLPBGITYHTTSSYLLHDDGRAVVTYEGDTSAGY	NHDVSLTRAM	IRTTKIVGDH	240						
D	b	190	lhlhpvllppgytchtsyllindgdraavlytgdsagjnhv	slrlrawitck	avggh	249						
Q	y	241	PLVIERRALGCHVLLTLTAAPESPMYPYPRSR	REVVYRSIFGGS	SPSLFPSACSRK	300						
D	b	250	plvieraralgchvllltaapepsmypyprsr	teevyrsifggs	spslfpsacsck	309						
Q	y	301	STFHAVENHIDWRLMLFGATLIDAOAFCCSRKMLTYL	LGISYKTVGALVANEC	MNASEDL	360						
D	b	310	stfhavrvhwdrlmlfgatlddgafccsrmltylgisyk	tvgalvanegmas	eddl	369						
Q	y	361	TAXITAAVLTICHORVLYRTQALISKMR	LGEHAOKITRLYS	WLFKSGRDYIPGR	420						
D	b	370	tavitaayltichbqyrlrtqaiskmr	lvehaqfictrlyswlf	feksgrdyipgr	429						
Q	y	421	FYAQCRWLSGPHLDRLRVLPFDESVCRCR	TLLKVAKGFCFCFMWLO	OBECTFLERAE	480						
D	b	430	fyagcrrwlsagphldrvlpfdesvrcrrtllkvagk	fcfcfmwlygectctle	epae	489						
Q	y	481	GLVGDHCHDNMAEGSEVDAPEABHLDVSTGYV	HHCHOLEALYRALNV	PODIAARASRLT	540						
D	b	490	glvgdyghdnmaegsevdapeabhlldvsgtyv	hghrqlaalylralnvphdla	aaasrlt	549						
Q	y	541	ATVELVASPDLRECTVLGKNTFR	TTVVGDANHL	EAANGPEQYVLSFDAS	RQSMGAGSHSLT	600					
D	b	550	atveltaspdlectrvlgkntfr	ttvvgdahnleangpeqyvl	sfdasrqsmgagsh	609						
Q	y	601	YELPRAGIQRWISNGIDJCATFRPPGAP	SAAGBYAACSALYRNR	TQOHSLSLGLM	660						
D	b	610	yelpraglqwrissngidcatfrppgapa	agbyaacsalyrnrtrqohs	slslglmw	669						
Q	y	661	LHPGGLGIFPPEPSGHIWESANPFC	EGTGLYRTWSTGSGSD	SEFPAPAA	PMAAATPG	720					
D	b	670	lhpgglgifpfpespghiwepanpfc	egtglyrtwtstgsgsf	sdsefpaa	aarpylaagp	729					
Q	y	721	LPBSTPVSVDIWLPRPSEEPQYDA	VPVPAPDAGLPG	CVLTL-PPRPV	NHKSIRPP	779					
D	b	730	lpbtpvsydlwlprrpsksyqdaasv	ppapapaglpjs	slvtlprpr	prtkprrpp	789					
Q	y	780	SRNRRLIYTPDDAKVYVAGSLFES	DCDMLYANSPCHRG	GGGICIAFYOR	FRFEPATYPTFF	839					
D	b	790	srtrlllytvpdgakvyagslfsedcmw	lynaenpghrpgg	lchalygrf	peaelypte	849					
Q	y	840	IMREGIAAUYTLTPRIIHAAPDY	RYEONKRL	LEAAYRECS	RGRGAAYPLLG	SGIYOAP	899				
D	b	850	imreglaaytltrpriihavapdyryeqn	pkrleaa	yretcstrg	taaprllygs	siyqvp	909				
Q	y	900	VSLSFDAWERNHREDELVLTL	ERPAANWF	EANKPAQ	QEVLTITTEDTA	TANLAL	LEIDATEVY	959			
D	b	910	vsisfdawernhrgdelvltlterpaanw	f	ean	kpaqgalctited	at	tanlaleida	tdv	969		
Q	y	960	GRAAGCTISP	GIYHYOTFAGVGS	GSKSRSTOOG	DVDVYVVP	PRREL	RNSMR	KRRG	GAATP	1019	
D	b	970	graaagctispglyvhygftagvpgs	ksrsisqgdv	dvvvvpltr	elms	wrrg	ftaafp	1029			
Q	y	1020	HTARAVITGRVVYVDEAP	SLPRHLLLLHMO	RASVHLDD	PMOPIA	IDENH	GLVPAIRP	1079			
D	b	1030	htaarvritgrvvyldeapslprhllllhm	grasvhlldpmpoia	idenhglvpa	irp	1089					
Q	y	1080	ELAPTSWMXYTHRC	PADVCELIGAYPK	IOJTS	SRVLSLF	WNEP	AI	GO	KLYVTO	QAQKAN	1139
D	b	1090	elaptswmxythrcpadvceligaypk	jtsrvtlsf	wnepa	igklyftg	qaakan	1149				
Q	y	1140	PGAITVHEAOGATF	ETTTATADAGL	IOS	FAHAIVAL	TRHTE	KVCYL	LD	AP	GLLEVG	1199
D	b	1150	pgaitvheagatfetttata	daglylqss	rahaivaltrh	tekc	vylladp	qjllrev	1209			

QY 1200 ISDVIVNNEFLAGGEVGHKPSVIRPGNDQNLGTLQAPSPSCQISAVYHQLAEELGHRPA 1259
 |||||||
 Db 1210 lsdvlnvfflaggevgnhhrpsvrlprgnpddqnlgtlqatfpescqlsaiyqhlaeelghrpa 1269
 QY 1260 PVAATLPCPELEHOGELTWPOLYSDSVLVEFLDVIYHCRMAASQKRAVSTLVGRG 1319
 |||||||
 Db 1270 pvaavllppclelqgllympgelvdsdsvlvelcdvncrmaapsqrkavistlvgrg 1329
 QY 1320 RRTKLYEAHSDVRESLARFIPTIGPVQATCELYELVAMYEKGODSASVLELDCND 1379
 |||||||
 Db 1330 rrtklyeaahsdvreslarfipitlipyvqatcellyelvamyekegddgsavleldcnd 1389
 QY 1380 VSRITFPQKXCKFTTGETIAHGKVGQGISANSTKFCALFGWPAIEKEIILPPNIF 1439
 |||||||
 Db 1390 varitffqkcnkfttgetiahgkvqgisawsktfcalfgwfrakeleiallppnlf 1449
 QY 1440 YGDAYEESVFAAAGAGSCMFENDESPDSTONNFIIGLECYVMEECGMQWILRLX 1499
 |||||||
 Db 1450 ygdayeessvfaaavagagscmfendfsefdstqnfisiglecyvmeecgmqwlrlx 1509
 QY 1500 LVRSAMIIQAPKESLKGEMKHSGEPGTLIMNTVMNMALIAHCYERDFRVAAFKGDSDV 1559
 |||||||
 Db 1510 lvrsawllqapkeslkgfkkhsgpgtllmwtvmnmaliahcyetrdfrvaafkgsdsv 1569
 QY 1560 VICSDFRQSRMAALIAAGCLKLYDYPRTGLYAGVYVAPGILTDPVVRFRGRUSEKMW 1619
 |||||||
 Db 1570 vicsdfyrqsrmaaalagcglklydyrptgllyagvvyvapgiltldpvyrfagrusekmw 1629
 QY 1620 GGPERRAEOLRLACDEFLRGLFNVAQVCDVVSRYGVSPGIVHNLICMLQITIAOGKAHF 1679
 |||||||
 Db 1630 ggperraeeqlrlavcdfirglfnvaqvcdvvsrygvspgvlhmligmlqladgkahl 1689
 QY 1680 TETIKPVLDTNIIQORVE 1698
 |||||||
 Db 1690 tetikpvlidtnsiigrve 1708

- RESULT 3

AAY31384
 ID AAY31384 standard; Protein; 1708 AA.

XX AAY31384;

XX 12-OCT-1999 (first entry)

DE HEV-US2 ORF1 protein.

XX Hepatitis E virus; HEV, binding partner; virus; US-HEV infection;

KW vaccine; passive immunisation.

XX Hepatitis E virus.

FT Key location/Qualifiers

FT Misc-difference 322 /label= unknown

FT Misc-difference 331 /note= "encoded by CNG"

FT Misc-difference 331 /label= unknown

FT Misc-difference 445 /note= "encoded by RAC"

FT Misc-difference 448 /label= unknown

FT Misc-difference 634 /note= "encoded by RAC"

FT Misc-difference 634 /label= unknown

FT Misc-difference 646 /note= "encoded by CCY"

FT Misc-difference 811 /label= unknown

FT Misc-difference 811 /note= "encoded by GCS"

FT /label= unknown
 FT /note= "encoded by TRF"
 FT Misc-difference 1533 /label= unknown
 FT /note= "encoded by TAY"
 FT Misc-difference 1578 /label= unknown
 FT /note= "encoded by RGC"
 FT Misc-difference 1691 /label= unknown
 FT /note= "encoded by GAR"

W09919732-A1.

22-APR-1999.

15-OCT-1998; 98MO-US21941.

15-OCT-1997; 97US-0061199.

(ABBO) ABBOTT LAB.

Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;

WPI: 1999-288017/24..

N-PSDB; AA200267.

Detection of United States isolates of hepatitis E virus
 Claim 6; Page 216-221; 260pp; English.

The invention provides a method for detecting a US (sub)type hepatitis E virus (US-HEV), or its naturally occurring variants in a sample by treatment with a binding partner specific for a marker of the virus, and then detecting any complex formed. The method is used to diagnose infection with US-HEV. Polypeptides from US-HEV, antibodies specific for open reading frames (ORF) in US-HEV and host cells expressing these ORFs are useful in vaccines or for passive immunisation. The polypeptides are also used to raise specific antibodies (useful as immunoassay reagents). CC Fragments of nucleic acid from US-HEV are useful as primers and probes in usual hybridisation and amplification assays for detecting infection. The present sequence represents a HEV-US2 ORF1 protein.

Sequence 1708 AA;

Query Match 97.1%; Score 8753.5; DB 20; Length 1708;

Best Local Similarity 97.2%; Pred. No. 0; Mismatches 39; Indels 1; Gaps 1;

Matches 1651; Conservative 8;

QY 1 PGTTAIEQALAAANSALANAVVVRPFLSRVQTEILINLMQPRQLVFRPEVLMNHPIOR 60
 |||||||
 Db 10 pgttaieqaalaaansalanavvvrpflsrvtellnmlmqprqlvfrpevlmnhpiqr 69
 QY 61 VTHNELDYTCRARGCLEVGAHPRSINDPNVLAHRCFPRGQVQRYASPTGSPAN 120
 |||||||
 Db 70 vthnledytcragrclevgaahprsinpnvlarhcfprgqvqrvysaptgspaan 129
 QY 121 CRRSALRGLPADRYTCFPGFSRCAFAAFTGVALISLHDLPADVAEAMARXTRLVA 180
 |||||||
 Db 130 crrsalrglpadrytcfgfsgsrcafaaftgvalyslhdlpadvaeamarxtrlvaa 189
 QY 181 LHLPREVLLPGCTHTTSTYLLIHDDDRAVVTEGDTASAGYNHDSILRAMIFRTKIVGDH 240
 |||||||
 Db 190 lhlprevlppgcthtntsllylhdgnravvtegdtsagynhdsvilrtwrttkivgth 249
 QY 241 PLVIERVRAIGCHEVLLTLADPESPMPVVPYRSTEVVRSIFGPGSGSPSLPSACSTK 300
 |||||||
 Db 250 plviervraigchfvllltaapepsmpvypyrstevvrsifpgsgspslfpaacstk 309
 QY 301 STEHANVPYHIMRMLFQATLDDQAFCCSRMLTYLRGTSKYKTVGALVANEGMNSEDAL 360
 |||||||
 Db 310 sthanvpynhwdxlmfqtalxdqafccsrmltylrgtskykvtvgalvanegwnsedal 369


```

OY 361 TAXITAAVLTICHOVRLTQATISKGRRLGVEHQAOKFITRLYSWLEPKSGROYIPRGQ 420
  |||||||
Db 370 favltaayltlchqyrlrqaalskgmrlevehaqflicrllyswlfeksgrdyiprglq 429
OY 421 FYAOCRRMLASAGFHLDPRLVFEDESVPCHRTFLKKVAGKFCCEFMWLOCECCELEPPE 480
  |||||||
Db 430 fvaqctrwlsagfhlxprxlivfdesvpcrcrflkkvgaqkccfmwlgqectcflepe 489
OY 481 GLVGDHGDNEAYEGSEVDPAPAPHLDVSGTYAVHGHOLEALYRALNVPDODIAARASRLT 540
  |||||||
Db 490 glvgdghdneayegsevdpapapahldvsgtyavhghlealyrainlvphdiaarasrlt 549
OY 541 AAVELVASDRLCECRVTYGNKTRFTVVGCAHLHEANGRPQOYVLSFPAASQSMAGSHSLT 600
  |||||||
Db 550 avelvaspdrlectvtynktrftvtvgahleangpreeyvlsfpaasqsmagshslt 609
OY 601 YELTPAGLQVRISSNGLDQTATFPFGAPSAAPGEVAAFCSALYRYNRTFORSLTGLW 660
  |||||||
Db 610 yeltpagqlvksinsngldctatfpgpapsaapgevaxacsalyrynrftqrhsltgqlw 669
OY 661 LHPEGLLGIFPPSPGHIWESANPECGEGTLVTRWSTSGFSDFSPPEAAPAMAAATPG 720
  |||||||
Db 670 lhpegllygifppspghiwesanpfcgegltlyrtwstsgfsdfsppeaapasaapq 729
OY 721 LPHSTPPVSDIWLVPSPSEFOVDAAPVPAPDPAGLPGPVIT - PPPPPVHKPSIRPP 779
  ||:|||||
Db 730 lphstppvsdlwlvppspseeshvdaasvpspepagltsplyltpppppvprkpslpp 789
OY 780 SHNRRLLYTPDGAKYVASLFEESDCMLVNASNPGHRRPGGICHAFFOYRFPAPFPTPE 839
  |||||||
Db 790 ptrlrllyrpgdakyvaslkesdcdwlynasnpghrpggqlchafyrtfpeafstef 849
OY 840 IMREGIAATLTPRPITIAVADRYEONPKRLAAAYRETCSRGTAAVPLLGSGIYOVP 899
  |||||||
Db 850 imreglaaytlprplihavapdyveqnpkrleaayretcsrgtaaypllgsglyyvp 909
OY 900 VELSPAMERNHRPGDELYLTERPAMWFEANKPACPVLTTEPTATNMLEIDATEY 959
  |||||||
Db 910 velstidawernhrpgdelyletpaawifeankpaccpaltfetedatnalleidatey 969
OY 960 GRACAGCTTSPGIVHKOYFAGVPGSGKSRSIOGDVGVVVPTRRLNSMRRRGFAATP 1019
  |||||||
Db 970 gracagcttspgivyhkyfagvpgsgksrsiqgddvrvvptrelrnsrrtgrfaatp 1029
OY 1020 HTPAARVYIGRRVVIDEAPSLPHLLLLHMQRASSVHLGLDPMQIPAIIDEHAGLVPAIRP 1079
  |||||||
Db 1030 htpaarvlygrvvideapslphllllhmqrassvhlldgpmqipaidehaglvpaairp 1089
OY 1080 ELAPTSMWKVTYHRCPADVCELIRGAVPKTQITSRVLRSLFWNEPAIGQKLYXTOAKAAN 1139
  |||||||
Db 1090 elaptsmwvthytrcpadvcelirgavpkqtsrvtlrslyfnepaigqklyvfqaakaan 1149
OY 1140 PGATVHEHOGATFTEETTLIATADARGLIOSSRAHAIVALTRHTECVLLDAPGLREVG 1199
  |||||||
Db 1150 pgaivhehogaftfettliatadargliqssrahaivaltrhtecvllldagplirevg 1209
OY 1200 ISDVIVNNEFLIAGGEVGHRRPSVIRGNPDONLGLTQAPPPSCQISAYSQLOAEELGHRPA 1259
  |||||||
Db 1210 isdvivnnefliaggevgghrrpsvirgpnrdnlgltqapppscqisaynglaeelghrrpa 1269
OY 1260 PVAAVIPLPELEOGILLMPDELIVSDSVLVELLDIVICRMAAPSQRKAVSTLVGRYG 1319
  |||||||
Db 1270 pvaaviplpeleogillympdelivsdsvlvelldivnormaapqkravstlvgyrg 1329
OY 1320 RRTKLYEAHSDVRESLAFRTPTIGVQATCELYELVAMVEKGGDSAVLELDICND 1379
  |||||||
Db 1330 rrtklyeaansdvreslartptigvratcellyelvamvekqgdsavleldicnd 1389
OY 1380 VSRITFEOKXCKFTTGETIAGKVGQGISIAMSKEFCALFGPWFRAIEEIIIALDPNTE 1439
  |||||||
Db 1390 vsritffgdkcnkfttgetiagkvqgisaisavsktcalfgpwfraieeiiiallpnlf 1449

```

```

OY 1440 YGDAYEESVFAAASGAGSCMVFENDPSEFDSQNNPISGLGECVMEECGMQWILIRYH 1499
  |||||||
Db 1450 ygdayeessvfaaavsgagscmvfendpsefscqnnfslglecvmeecgmpwllirlyh 1509
OY 1500 IYRSAMILQAPKESLKGPMKHHSEEGTLLMNTVMMAITAHCEYRDFRVAAFKDDSV 1559
  |||||||
Db 1510 lyrsamilqapkeslkgpmkhhseegtllwntvmmaitahceyrfdrvaafkddsv 1569
OY 1560 VILCSDYRQSRNMAALJAGCGLIKLVDRPIGLIYAGVVAPEGLTLPDYVFAGRLEKMW 1619
  |||||||
Db 1570 vilcsdyrxrmaaaljagcgliklvdyrpiglyagvvapegltlpdyvfaagrlseknw 1629
OY 1620 GPGPERAEQURLAVCDPLNGITNVAQVYVSRYVGSGLVHNLIGMLQTTADGKAHF 1679
  |||||||
Db 1630 gpgperaegurlavcdfplngitnvayvysryvgsylvhnligmqlttadgkahl 1689
OY 1680 TETIKPVLDTNSIIORVE 1698
  |||||||
Db 1690 tnikpvlcltntsiigrve 1708

RESULT 4
AAR91813
ID AAR91813 standard; Protein; 1693 AA.
XX
AC AAR91813;
XX
DT 26-NOV-1996 (first entry)
XX
DE Hepatitis E virus strain SAR-55 ORF-1.
XX
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
KW structural region; antigen; detection; antibody; vaccine;
KW immunisation; infection.
XX
OS Hepatitis E virus.
XX
FH key location/Qualifiers
FT Misc-difference 1238 /note="corresponding codon CAG"
FT FT
FT Misc-difference 1455..1693
FT /note="10 bp nucleic acid sequence TGCTNTTGYCA
has to be inserted between nucleotides
4390..4391 of AAT27394 before these amino
acid residues can be decoded"
FT FT
FT MO9610580-A2.
XX
PD 11-APR-1996;
XX
XX 03-OCT-1995; 95WO-US13102.
XX PF
XX PR 03-OCT-1994; 94US-0316765.
XX PA
XX PI (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PT Emerson SU, Purcell RH, Tsarev SA;
XX PR WPI: 1996-209320/21.
XX PS N-PSDB; AAT27394.
XX
XX Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
PT antigenic protein useful in diagnosis, prophylaxis and treatment of
PT hepatitis E virus infection
XX
XX Disclosure; Pages 9-13; 121pp; English.
XX
XX The present sequence is the protein prod. of ORF-1 from the
CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an
CC enterically transmitted non-A, non-B hepatitis in Pakistan. The
CC protein encoded by the structural region of the virus (i.e. ORF-2),
CC which is capable of forming HEV like particles, is useful for the
CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,

```

CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The
 CC protein, and anti-HEV antibodies generated using the protein, can
 also be used in vaccines for immunising an animal against HEV
 CC infection. The protein is identified as a band of greater than
 CC 50 kD following SDS-PAGE of cell lysates of insect cells infected
 CC with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
 CC expression vectors pPIC9-1779, -1780 and -1781.

XX Sequence 1693 AA;

Query Match 82.7%; Score 7456; DB 17; Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 1400; Conservative 110; Mismatches 160; Indels 42; Gaps 7;

QY 1 PGITTAIDQALAAANSALANAVVRPELSRVQTELLIMLPROLVPEVLMNPIOR 60
 DB 10 pgittdaqaalaanaaanaavvrpflshqqllellimqprqlvtrpevfwhpqlqr 69
 QY 61 VTHNELEQCRARAGRCLEVGAPRSINDPNVLRHRCFLRPVGRDQORVYSAPTGPAAN 120
 DB 70 vthneleqcraragrclevgaprsindpnvlnrcflrpgdvqrvytabrtqpaan 129
 QY 121 CRRSALRGLPPADRTYCEGFSRCAPAAETGVALYSLHDLMPADVAEAMARSGXRLYAA 180
 DB 130 crrsalrglpaadrtycfdgfsgrcnpaetgvalyslhdmrspdvaeamfrhgmtrlyaa 189
 QY 181 LHLPEVLLPCTYHTTSTVLTHDGDRAVYTESGTSAGYNNDVSLKAWITTKIVGCH 240
 DB 190 lhlpevllpctyltasyllthdgrvrvvlyegdtsagynhndvslkwtvttkvgtgh 249
 QY 241 PLVIRRVALGCFVLLTLAARPESPMPVYPVPRSTEVVRSIFPGSGSPSLFSPASCSTK 300
 DB 250 plvirrvvalgcfvlltlaarpepmpvypvprstevvrsifpgsgspslfpscscst 309
 QY 301 STEFAVVPVHIDRLMFGATLDDQAPCCSRMLTYLRGISYKVTGALVANEGSMASDAL 360
 DB 310 stfhnvaphlwdrlmfgatlddqaqfcsrlmtylrgisylkvtgclvanegwmasedal 369
 QY 361 TXHIAAYLTICHOHRLFRQALSKGRMRGLVGHAKETIRLSWLFEEKSGRITVGRORQ 420
 DB 370 txviaayltichohrlyltqalskgrmrlerehaqfllrlyswlfeksgridytpgrtle 429
 QY 421 FYAQCRRMLASGFHLDPRVLVDESVPCRCRTFLKRVAKFCCEFMKMLGOECTCELEPAE 480
 DB 430 fyaqcrmlasgfhldprvlvdesaprcrtaikrkvakfcfcfmkmlvgectcfiqpae 489
 QY 481 GLYGHGHDNEAYEGSEVDPAEPANLDVSGTYAVHGHOLEALYKALNPQDIAAARSKLT 540
 DB 490 glyghghdneayegsedvdpaeasalsdsgsyvvpqtalqplqgaldlpaeltvaagrlt 549
 QY 541 ATVEELVAPSDRLRCTVGNKTFRTTVVDGAHLLEANGPQVLYLPDASQSGAGSHSLT 600
 DB 550 atvaelvapsdrlrctvgnktrfttvvdgaahlleangpovlylpdasqsgagshslt 609
 QY 601 YELPAGLQVIRISSNGLDCTATFPFGGASAPAAEVAACALYKXNRPRTQHNLSLTGLIM 660
 DB 610 yaesaglvirissngldctatfpfggasaapaaevaaacalyxnrrprrtqhnslstglim 669
 QY 661 LHPEGLGIFPPSPGHIWESANPFCGEGTLTFRIMS-----TSGFSSDFSPSPE 709
 DB 670 lhpegllgfpfpfpghiwesanpfcgestlylfrtswsevdavspapqdlgfts-----e 724
 QY 710 AAAPMAATPGLPSTPVSDIWLPRPSEERQVDAAPV-PPAPPA-GLRPPVULTLPPP 767
 DB 725 psiperaatp-----tp-----aaapppapdpstlspargarpp 761
 QY 768 PPVHKPSIP-PPSHNRRLLYTPDGAKYAGSLFESDCDMLVYNSNGHRRPAGGGLCHAF 826
 DB 762 gatarapltbqtarhrrllftypdgskvfagslfestctwlvnsanvdhrpaggglchaf 821
 QY 827 YORPEAFYPEEFIRREGILAAVTLTPRPIIAAVADRYEONPKRLAEAVRETGCRKGT 886
 DB 827 yorpeafypeefirregilaavtltprrpiiaavadyeonpkrlaeavretgcrkgt 886

DB 822 ygrtpasfdaasfvmrdgaaytlcprlphnavapdyrlenhpkrlaeaayretscrlgta 881
 QY 887 AYPILGSGTGYQVPVSLSPFAMERNHRPGDELVTFEPAAMWEPANPAQVULITTEDTART 946
 DB 882 aypilgsgtgyvpvslspfaernhrpgdelvtfepaamwepanpaoqvulittdedart 941
 QY 947 ANIALEIDATVEGRCACGCTTSPGVHVOYFAGVPGSGSKSSIOGDVVVVVPTRELR 1006
 DB 942 anialeidatvegrcacgcttspgvhvoayfagvpgsgskssioogdvvvvptrelr 1001
 QY 1007 NSWRRRGFAAPTPHTAARTYITRRVVIDEAPSLPHLLILHMQRASSVHLLDPPQIPAI 1066
 DB 1002 nswrrrgfaaptphtaartyttrrvvideapslphllilhmqratvhlldppqipa 1061
 QY 1067 DEHAGLVPAIRPELAPTSMMWVTHRCPADVCELRGAVPKIQTMSRVLSLFWMPEPAIG 1126
 DB 1062 dehaglvpaairpelaptsmmwvthrcpadvcelrgavpkiqtmsrvlsfwmpepag 1121
 QY 1127 OKLVYTOAKAANPAATVHEAQAFTTETTLIATADAGLTIQSSRAHAIVALTREHTEKC 1186
 DB 1122 oklvftgaakaanpaatvheaqafettetlliatadagltiqssrahaivaltrehtekc 1181
 QY 1187 VILDAAGLIREVIGSDVIVNFFLAGGEVGHKRPVYIPRGNDQNLGTLQAFPPSCQISA 1246
 DB 1182 vildagpdlirevigsdvinyvfflaggevgkhkrsyviprgndqnlgtlqafppscqlsa 1241
 QY 1247 YHQLAEELGHRPAPAAVYPPCPBELGGLVYMOELTVSDSLVLELMDIVHCRMAAPSQ 1306
 DB 1242 yhqlaeelghrpapaaavyppcpbelgglvymoeeltvdsdsvlelmdivhcrmaap 1301
 QY 1307 RKAVALSTLVGRYGRRTKLYEAAHSDVRESLAFPIITGVPVQATTCELYELVEAWVEKGD 1366
 DB 1302 rkavaltstlvgryrtrtklynaashdvsdrlarfipaigvqvttcelyelveawvek 1361
 QY 1367 GSAVLELDICNDVSRITFPQKXCKNFTTGETIANGKVGQGISANSKFTCALFGWFRAI 1426
 DB 1362 gsavleldicndvsrirtfpqkxcknfttgetiangkvqgisanskftcalfgwfrai 1421
 QY 1427 EKEIITALPPNIFYGDAAVEESFYAAVSGAGSCWFEFSEFSDTQNNFSLGLBCVYME 1486
 DB 1422 ekeitalppnifygdaveesfyaaavsgagscwfeffsefsefsgtqnnfslglbcvme 1481
 QY 1487 ECGMPQWLIRLYHLVRSAMITLQAPKESLKGFMKKHSGEPGILLMNTVMNAIILANCYER 1546
 DB 1482 ecgmpqwlirlyhlvrsamitlqapkesslkgfmkksgepgillmntvmnailancyer 1541
 QY 1547 DFRVAAFPGSDSVVLCSDYROSRNAALILAGGGLKLVXDYRPIGLYAGVVVAPGILTPD 1606
 DB 1542 dfrvafpfgsdsvvllcsdyrosrnaalilaggglklvxdyrpigliyagvvvapgilt 1601
 QY 1607 VYRFAGRLSEKMMWGPGERAEQRLAVCDFLGLTNNVAQVDVVSRYVGYSPGELVHNL 1666
 DB 1602 vyrfagrlsekmmwgpgeraeqrlavcdfllgltnnvaqvdyvvsryvgyspgelvhnl 1661
 QY 1667 GMLQTIADGKAHFTETIKPVLDLINSIIQRYE 1698
 DB 1662 gmlqvadgkahnfteskvpxldlinsilcve 1693

RESULT 5

AAW80196 standard; Protein: 1693 AA.

AAW80196;

23-DEC-1998 (first entry)

Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.

Enterically transmitted non A non B hepatitis virus; ET-NANB;

Hepatitis E virus; HEV; Burma HEV isolate; vaccine;

diagnostic probe.

OS Non A non B Hepatitis virus.
XX US5824649-A.
XX 20-OCT-1998.
XX
XX 07-JUN-1995; 95US-0475807.
XX
XX 25-JUL-1994; 94US-0279823.
PR 17-JUN-1988; 88US-0208997.
PR 11-APR-1983; 89US-0336672.
PR 16-JUN-1989; 89US-0367486.
PR 13-OCT-1989; 89US-0420921.
PR 05-JUL-1990; 90US-0505888.
PR 07-JUN-1995; 95US-0475807.
XX
XX (GENE-) GENELABS TECHNOLOGIES INC.
PI Bradley DM, FTY KE, Krawczynski KZ, Reyes GR, Tam A;
PI Yarough PO;
XX WPI: 1998-582589/49.
DR N-PSDB: AAV65321.
XX
PT Hepatitis E virus proteins - useful for diagnosis or vaccine
production the virus
XX
XX Claim 22: Columns 57-66; 47pp; English.
XX
XX AAW80196-98 are encoded by the genome of the Burma strain of
- CC enterically transmitted non A non B hepatitis virus (ET-NANB)
CC (hepatitis E virus (HEV)). The specification describes an isolated
CC protein which is specifically immunoreactive with antibodies present
CC in individuals infected with HEV and encoded by a sequence contained
CC in an open reading frame (ORF) of an HEV genome. The genome has a
CC sequence that is more than 70% identical to the ORF1 sequence from
CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic
CC probe for ET-NANB.
XX
XX Sequence 1693 AA:
SQ

Query Match 82.6%; Score 7448; DB 19; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY 1 PGITTAIEQALAAANSALANAVVVRPFLSRVOTELINIMOPROLVFRPEVLMNHPIQR 60
DB 10 pglttalegaalaanaalanavvrplshqgtellinlmprqlvfrpevtmnhpigr 69
QY 61 VTHNELQYCRARAGRCLEVGAAHPRSTINDPNVLRHCFLRPVGRDVORWYSAPTRGPAAN 120
DB 70 vthnelqyrcarsgrclcgahprstindpnvhrchflrpvggrdvgrwytaptrgpaan 129
QY 121 CRRSARLGGPAPARTYCOFSESCAFEAERGVVALYSLHDMAPADVAEAMARHGXTRXYAA 180
DB 130 crrsarlgpapartycldgfcgcnlpaetclgialyslhmsspdvaeamtrhgmtrlyaa 189
QY 181 LHLPEVLLPFGYHTTSYLLIHNDGRAVVTYEGDTSAGYNHDVSIIRAWIRTTKTYGDH 240
DB 190 lhlpevllppgtyctasyllihndgrrvvtlyegdtsagynhdvsnlrsirttkvqgh 249
QY 241 PLVIERVRAIGHFVLLTLAAPPSPMPVYVPRSTEVVRSIFGPGCSPLSPSACSTK 300
DB 250 plviervraigfhfvllltaapepsmpyvpvrstevvrsifpgqgtpslfpstscstk 309
QY 301 STEHAPVPHWDRMLFGATLDDOAFCCSRLMTYLGISYKVTGALVAGVANEAGNNSDAL 360
DB 310 stehapvphwdrmlfgatlddofccsrlmtylgisykvtgaltvaneagvnsadal 369
QY 361 TAAITAAVLTICHQRYLRTQAIKSGMRRLGVEHAOKFTITLYLWLFKSGRDYTPGRLO 420
DB 370 tavitaaayltichqrylrgaiskmgmrllgvehaokftitlylwlflksgrdytpgrgle 429

QY 421 FYAQCRRLWLSAGFHLDPRLVFPDESVPRCRRTFLKKAAGKFCCEPMRLAGECTCLEPAE 480
DB 430 fyacqrwlwsagfhlldprlvfpdesvpcrrtflkkaagkfccepmrlagectclepae 489
QY 481 GLVGDHNDNEAEGSEVDPAEPANLIDVSGTYAVHGHQLEALYALNVPDIDARASRLT 540
DB 490 gavgddghndneavgsevdpaepanlidvsgtyavhghqlealylalnvpdidaarsrlt 549
QY 541 ATVELYASPDRLRCRYVLAKKTRTTYVDGAHLBANGPEQYVUSFDSKOSMGASHSLT 600
DB 550 atvelyaspdrlrcryvlakkttrttyvdgaahlbangpeqyvusfdskosmgashslt 609
QY 601 YELPLPGLOYRISNGLDCATFPPGAPSAAPGEVAFCSALYRYRFRORSHLTGWL 660
DB 610 yaasaelvryyvaagldhavlapyvprsapgevalfcsalylrtrreagrhltylgnlw 669
QY 661 LHPEGLGIFPPSPGHIWESANPFCGEGTLYRTWS-TSGFSSDFSP-----PEAAPA 714
DB 670 fhpegllgltfaptspgjhwesanpfcgegtlyrtwseavspapdlgfmseps 729
QY 715 MAATPGLPHSTPVSIDIWLPSPSEEFQVDAAPV-PPAPDPAGLP-GRVVLTPPPPPVH 772
DB 730 raatpfl-----aatpppdpdpppsapaalaepasgtag 766
QY 773 KPSIP-PPSRNRRLTYTPPGAKVYAGSLFESDCDMLVNSNGHRGGGLCHAFYGRF 831
DB 767 apaltqltarhtrllftypgskvtagstestctvlnasnvdhprpgglchafygrp 826
QY 832 EAFVPTFEIRREGIATLTPRPIIHAVAPDYREONPKRLEAAYRTGCRBGAAAPL 891
DB 827 asfdaasfmrtdgaatllcpriinahapolytlempkrlaayetecsrlygtaapll 886
QY 892 GSGIYQVVALSPDAMERNRPPGDELYLTPANWPEANRPAOPVLITTEDTANTANLAL 951
DB 887 gtglyqypisrtdawernhrppgdelyltpelanwpeanrptpvlitedvartanla 946
QY 952 EIDAAETEVGRACAGCTTSPGIVHQTFTAGYVPGSKSKSIQGGVDVYVYVTRRLNSMR 1011
DB 947 eldsaltvgracagctrvtpvgyqftagvpgsksksriscgadadvvvpblrelmawr 1006
QY 1012 RGFAPATPHTAARTIGRRVVIDEAPSLPHLLLLHMQARASSVYLLGDPOIPADPEHA 1071
DB 1007 rgfaatphtaatrtgrrvvideapslphllllmqaratvhllygdpoipadideha 1066
QY 1072 GLVPALRPBELAPTSWMAVYTHRCPADVCELRGAYPKIQTTSRYLRSFLWNEPAIGQRLVX 1131
DB 1067 glvpalrpdtpgtswmavthtrwpadvcelrgraypmqtsrlyrsflwgepavqglvf 1126
QY 1132 TOAKKANPGALIVHEAOGATFTETTTIATADARGLIOSSRAAIVALTTRTEKCVTLDA 1191
DB 1127 tgaakpangpavltvheagatftettliatadargliqssrahaivaltrhtekcvilda 1186
QY 1192 PGLIREGISDVYVNNFFLAGGEVGHXRPVYIPGNDONIGTQAPPPSGQISAYHQLA 1251
DB 1187 pgliregisdvynnfllaggevgxhrrpvyipgndonigtqapppsgqisayhqla 1246
QY 1252 EELIGRPAPVAAVLPCCPELEOGILYMPQELVTYSDSLVVELLDIVICRAAPASQORAVL 1311
DB 1247 eelighrpvvaavlppeleogillypgealtcdsvvtecltclivncmaapqsrav 1306
QY 1312 STLVGRYGRRTKLYEAASDVRESLARFPTIGVQATTELELYEVANVEKODGSAYL 1371
DB 1307 stlvgrygrrtklyeaasdvreslarfptigvqattelelyevanvekgodgsayl 1366
QY 1372 ELIDCNRDVSRIITFEOKXCKFTTGENTIAHGKVGOGISANSKTPCALFGWFRATEEIL 1431
DB 1367 elidcnrdvsritftgdcckfttgetcliangkvvggisasktpcalfgwfrateail 1426
QY 1432 ALPPLNIFGDAYDEESYFAAASVAGSAGSCMFENDFSEFSDTQNNFSLGLCVVMEEGCMP 1491
DB 1427 allpplnifgdaydeesyfaaasvagsagscmfendfsefstqnnfslglcalmeegmp 1486

QY	1492	OMITELHYLVNSAWTILQAPKRSIKSGPKKHSGEPTLLAMNTVMMAITIAACIEPRDERFA	1551
Db	1487	qmlitrltynhltsawlllqapkesltlrgfwkhsqgeplllwntvmavltlcydrtdrtqva	1546
OY	1552	AFKGGDSVVLSDYKQSNMAALITAGCGIKAKVDYRPIGLYAGVVAVPGJGTTLPDVRFA	1611
Db	1547	afkgdssvlscseyqspgaavllagcglikvdfirpigliyagvvaavgjgaldpvrfa	1606
OY	1612	GRLSKKNNGPGEPRAEQLRLAVCDPLRGJINVAOCVDVVSRYVGSFGLVHNLIGMLQRT	1671
Db	1607	grltekknwpgpaeaeqrlrlavsdfflxfkltnvagmcvdvsryvgsfplvhnllgmqlga	1666
OY	1672	TADGKAHFTETIKPVLDTLNSIIQRYE	1698
Db	1667	vadgkahftesvkpyldltnslcfrve	1693
RESULT 6			
AAB24119			
XX	AAB24119 standard; Protein: 1693 AA.		
AC	AAB24119;		
XX			
DT	29-JAN-2001 (first entry)		
XX			
DE	Hepatitis E virus Burma strain ORF1 protein sequence SEQ ID NO:7.		
XX			
KM	Hepatitis E virus; HEV; non-A/non-B viral hepatitis; viral protein;		
KW	immunological; diagnosis; hepatitis; infection; identification;		
KM	detection; immunoreactive; hepatotropic; antiinflammatory; virucide;		
KW	vaccine; antiviral; antigenic; antibody; antigen.		
XX			
OS	Hepatitis E virus.		
XX			
PN	US6120988-A.		
XX			
PD	19-SEP-2000.		
XX			
PE	07-JUN-1995; 95US-0478507.		
XX			
PR	25-JUL-1994; 94US-0279823.		
PR	05-APR-1991; 91US-0681078.		
PR	17-JUN-1988; 88US-0208997.		
PR	11-APR-1989; 89US-0336672.		
PR	16-JUN-1989; 89US-0367486.		
PR	13-OCT-1989; 89US-0420921.		
PR	05-APR-1990; 90US-0505888.		
XX			
PA	(GENE-) GENELABS TECHNOLOGIES INC.		
PA	(USGO) US GOVERNMENT.		
PI	Yarborough PO, Krawczynski KZ, Fry KE, Bradley DW, Tam A, Reyes GR;		
XX			
DR	WPI: 2000-593713/56.		
XX	N-PSDB: AAA55259.		
XX			
PT	Identifying recombinant antigen immunoreactive with antibody induced by		
PT	hepatitis E virus (HEV), for detecting HEV infection, comprises		
PT	immunoreacting a polypeptide from an HEV genome with an HEV-positive		
XX	antiserum -		
XX			
PS	Claim 6; Column 55-64; 46pp; English.		
XX			
CC	The present invention describes a method for identifying a recombinant		
CC	antigen immunoreactive with a hepatitis E virus (HEV)-induced antibody.		
CC	The method comprises producing a polypeptide derived from an HEV genome,		
CC	immunoreacting the polypeptide with an HEV-positive antiserum and		
CC	selecting the polypeptide as a recombinant antigen if the polypeptide		
CC	reacts with the HEV-positive antiserum. The method is useful for		
CC	identifying recombinant antigen immunoreactive with antibody induced by		
CC	HEV. The enterically transmitted non-A/non-B hepatitis (ET-NANB (also		
CC	known as HEV))-specific fragments are useful for identifying ET-NANB-		
CC	derived cDNAs, which contain additional sequence information, as primers		

	for detecting ET-NANB viral genomic material in a patient sample, for	
CC	the synthesis of polypeptides for use in immunoassays, and for	
CC	identifying similar antigenic regions encoded by related viral strains,	
CC	e.g. Burmese strain. The antigens are especially useful in the	
CC	preparation of vaccine against ET-NANB infection. These antigens may	
CC	further be used to prepare antibodies to ET-NANB virus particles for	
CC	use directly as antiviral agents, and to produce antiserum designed for	
CC	pre- or post-exposure prophylaxis. The present sequence represents a	
CC	specifically claimed HEV Burma strain protein sequence for use in the	
CC	present invention.	
XX		
SQ	Sequence 1693 AA;	
	Query Match 82.6%; Score 7448; DB 21; Length 1693;	
	Best Local Similarity 81.8%; Pred. No. 0;	
	Matches 196; Conservative 110; Mismatches 169; Indels 32; Gaps	
QY	1 PGITTAIEBAALAAANSALANAVVRPLSRQTETLNLMPROLYPRPEVLMMHPOR	60
DB	10 pGItAlEgaaLaanaSaLaNaavvrPlShqJellInImPrJdlvIreVfwhPIdr	69
QY	61 VHNHEOXCRRARAGCEVGAPHSINDNPVWLHRCFLRPGRVQRMYSAPTGRPAAN	120
DB	70 vhnhelelycrrsgrclelgahprSIndnpvhncflrvgrldqvrytpeptgpaan	129
QY	121 CRRSALRGLPPADRTYCFDFGSRCFAFETGVALLSLHDLPADVAEAMRHGXTRYLAA	180
DB	130 crrsalrglpadrtcydIdfgscfnPaetGialySlhmSpdsvaeanfrngmtcrylaa	189
QY	181 LHLPREVLLPCTYHTTSYLILHHDDRAVVTREGTSGAGYNHDVSLTRWIKTKTVGDH	240
DB	190 lhlPrevlllprgcltyaSylllhDdrIrVVvvegdtsagynhdvsnlSwltKvtvgth	249
QY	241 PLVIERKRALIGCHFYLLTLAAREPSMPVVPVPRSTEVYVRSIFPGSGSPSFFASCTK	300
DB	250 plvieerVaigchfyllltaarepsmpvyprstevyrvislfpgpgspslfpcscsk	309
QY	301 STEFHAVPNHIWMRLFGATLDDQAFCSSRLMTYLRGISYKVTVGALVANEGMASEDA	360
DB	310 stefhavpnhiwrlmfgaTlddqaFcCsrLmtylrGisykvTvglVanegmaseda	369
QY	361 TAAITRAAYLTICHORYLKRQAISKMRRLGVNHAOKFTIRRLSWLFENSGRYTGROIQ	420
DB	370 taVitaaaylftIchorylkrQaiskmrtrerehaqkflTrlyswlfeksgrydIpgrgle	429
QY	421 FAACORRMVLSAGFNHLDPRVLVEDESIPCCSRFLFKLVAKFCFMKRMLODECTCFLEP	480
DB	430 fYaCOrtwlsagfnhldprvlvesapccrtaikalskfcmvmvjgectcfllgpee	489
QY	481 GLVGGHGHNENEYEGSEVDPAEPANLIDVSGTVAHGHODEALYRALNVQDIANAARSRLT	540
DB	490 glvggghgneneYegsevdPaesaelsldisgyvvvpgtalprlygaldlpraeltvaragt	549
QY	541 ATVELVASPEDRLERCTVUGNKTFRTTYVDGAHLNANGREPOYVLSFDASFOSKAGASHSLT	600
DB	550 atvksyvqdgridceotllgnktfftsfdvgaJlengeprhnlstfaasqtmaaarfsit	609
QY	601 YELTFRAGIOVRISNSNGCDCTATFRPGGAPSANAPGEVAFCSALYVKNFTQHSLTGLIM	660
DB	610 yaasaagleviryvaagldhravfagvsprspaevetafcsalylfnreaqrhslilgllw	669
QY	661 LHPEGGLGFPPFSGSHIWESANPCGDETLTYTRFWS--TSGSSSPFSP-----PEAAPA	714
DB	670 fhpeglllgfapfsghnwesannpcgeeltyltrtksewdavssparpdligtmsepshps	729
QY	715 MAATEGLPHSTPPVSDIWLPPRSEEFOVDAPV--PAPADPAGLP--GPVILTPPPPYPVN	772
DB	730 raatPltL-----aaPlppdpdppppsapaalaepasagatag	766
QY	773 KPSIP-PPSRNRRLLYTPBDCAKYTAGSLFESDCOMLVANASVPGRRPGCGLCNAFYQRP	831
DB	767 apathcarharhrliflycpdksvfagsfstecfcwlvanavdhirpaggclchaIfygxyr	826

OY	361	TATAATATTC	CHORRYRTA	ISAGMRIG	AEHAK	FEIRTR	YSWLF	EKS	GRD	Y	IGRO	420																																					
Db	370	tavlaa	yltcltq	ryltlq	aisgm	rleren	aqk	fltr	lyswl	eks	gdy	lpr	qle	429																																			
OY	421	FYAOC	RMLIS	AGFIH	DRPV	IVFDE	SVYCR	RTFLK	YVAG	FCF	FMWL	COE	CTCL	EPAE	480																																		
Db	430	fyaoc	rmlis	agfih	drpv	ivfde	svy	rcr	rtflk	yvag	fcf	fmwl	coect	cl	epae	489																																	
OY	481	GLVGD	HGHDNE	AYG	SEVD	PAEP	AHL	DV	SCT	YAV	HGH	QLE	ALY	RAL	N	POD	IA	RAS	LT	540																													
Db	490	gavdg	qndne	aysg	dvdp	aeas	ids	isg	svy	vp	tal	qly	gal	ld	prae	lva	rat	ag	lt	549																													
OY	541	ATVEL	YAS	PD	LE	ORT	Y	L	G	K	T	E	T	T	Y	V	O	S	A	H	E	A	N	G	P	O	Y	V	S	P	A	S	Q	S	A	C	H	S	H	T	600								
Db	550	atvel	yasp	del	ec	rt	yl	g	k	t	e	t	t	y	v	o	s	a	h	e	a	n	g	p	o	y	v	s	p	a	s	q	s	a	c	h	s	h	t	609									
OY	601	YELP	P	A	G	L	O	V	A	I	S	S	N	G	D	C	A	T	E	P	P	G	C	A	P	A	A	E	G	E	A	A	C	S	A	L	Y	R	N	R	T	O	H	S	T	G	L	M	660
Db	610	yael	p	a	g	l	o	v	a	i	s	s	n	g	d	c	a	t	e	p	p	g	c	a	p	a	a	e	g	e	a	a	c	s	a	l	y	r	n	r	t	o	h	s	t	g	l	m	669
OY	661	LHP	E	G	L	G	I	F	P	P	S	P	G	H	I	M	E	S	A	N	P	C	G	E	T	L	Y	R	T	W	S	-	T	S	P	S	D	E	S	-	---	P	E	A	A	P	A	714	
Db	670	lhpeg	l	i	g	f	a	p	e	g	h	i	m	e	s	a	n	p	c	g	e	t	l	y	r	t	w	s	-	t	s	p	s	d	e	s	-	---	p	e	a	a	p	a	729				
OY	715	MAATP	G	L	H	S	T	P	P	S	D	L	W	L	P	P	S	E	E	Q	V	N	A	A	V	-	P	P	A	D	P	A	G	L	-	G	P	V	L	T	P	P	P	P	P	Y	H	772	
Db	730	maatp	g	l	h	s	t	p	p	s	d	w	l	p	p	s	e	e	q	v	n	a	a	v	-	p	p	a	d	p	a	g	l	-	g	p	v	l	t	p	p	p	p	p	y	h	776		
OY	773	KPSIP	-	P	S	R	N	R	L	L	Y	T	P	G	A	K	Y	A	G	L	E	P	S	O	C	D	L	V	N	A	S	N	G	H	P	G	G	L	C	H	A	O	A	R	F	831			
Db	787	apalt	h	g	t	a	r	n	t	r	i	l	f	y	p	g	s	k	f	a	g	s	i	e	s	c	t	a	l	y	a	s	n	v	d	r	p	y	g	l	a	h	y	q	r	p	836		
OY	832	EAF	P	T	E	E	I	M	E	G	L	A	A	T	T	P	R	P	I	I	A	V	A	D	V	E	N	O	N	P	K	L	E	A	A	V	R	E	T	C	S	R	G	T	A	P	L	891	
Db	837	asfa	e	s	f	a	e	s	i	m	e	g	a	a	t	t	p	r	i	a	v	a	d	v	e	n	o	n	p	k	l	e	a	a	v	r	e	t	c	s	r	g	t	a	p	l	896		
OY	892	GSG	I	T	O	V	P	V	S	L	S	P	A	M	E	R	N	R	P	O	D	E	L	T	P	P	A	N	E	A	K	P	A	O	P	V	L	T	T	E	D	I	A	R	A	N	A	L	951
Db	887																																																

Accession	Source	Protein	Length	Weight	PI	Inst	Accession	Source	Protein	Length	Weight	PI	Inst
Db	1477	allpgrvtrfgdadtdvtrfsaaavaakaasmvfendfseifstqgnntslgdecaimecgmp	1486				Db	1487	gwlirlyhllrsaalldqapkeslrfgwkkhsgsgpflllwntvwmaavltbchcydfdrfdgva	1546			
Qy	1492	QMLIRLYHLVRSAMIILOAPRESLKGFWKKHSGSEPTGLLWNTVWMAIIAHCEERDFRVA	1551				Qy	1552	AFKDDSDVSLCSQVROSGRNAAALIIACGGLKIKYDYPICLTLYAGVVAVAGLGTLPPVYRFA	1611			
Db	1547	afkddsdvslcseqvrgspgaavllaagcgllklydrp19lyagvvavapglpdrvifa	1606				Qy	1612	GRLESEKMMGPGCEPAEQLRLAVCDPFLRGTLTNVNAQVVDVAVSRVYGVSPGLVHNLIGLOT	1671			
Db	1607	grlcekmgpgpcepaeeqrlrlavcdpflrlgtltnvnaqvmcdvavsrlygvspglvnhlignlga	1666				Qy	1672	IADGKAHPTETIKPVLDLNLSIIQRYE	1698			
Db	1667	vadgkahnfeevkpvldltlnslilcrve	1693				RESULT 8						
AAW81519		standard; Protein; 1693 AA.					AAW81519						
AAW81519		02-FEB-1999 (first entry)					AAW81519						
DE		Hepatitis E virus (HEV) ORF-1 protein.					DE						
XX		Hepatitis E virus; HEV; SAR-55: diagnostic agent; vaccine; antibody;					XX						
XX		passive immunisation; open reading frame; ORF.					XX						
XX		Hepatitis E virus.					XX						
XX		Hepatitis E virus.					XX						
XX		Key					XX						
XX		Location/Qualifiers					XX						
XX		Misc-difference 1238					XX						
XX		/note= "encoded by CAG"					XX						
XX		Misc-difference 1244					XX						
XX		/note= "encoded by CAG"					XX						
XX		Misc-difference 1352					XX						
XX		/note= "encoded by GTC"					XX						
XX		Misc-difference 1662					XX						
XX		/note= "encoded by GGC"					XX						
XX		WO9846761-A1.					XX						
XX		22-OCT-1998.					XX						
XX		09-APR-1998; 98WO-US07418.					XX						
XX		11-APR-1997; 97US-0840316.					XX						
XX		(USSH) US DEPT HEALTH & HUMAN SERVICES.					XX						
XX		Emerson SU, Purcell RH, Robinson RA, Tsarev SA;					XX						
XX		WPI: 1998-568733/48.					XX						
XX		N-PSDB; AAV17604.					XX						
XX		New hepatitis E virus DNA from Pakistani strain SAR-55 - used for,					XX						
XX		e.g. developing products for diagnosis of, and vaccination against					XX						
XX		hepatitis E virus infection					XX						
XX		Disclosure; Pages 11-15; 204pp; English.					XX						
XX		This represents a hepatitis E virus (HEV) open reading frame (ORF)-1					XX						
XX		protein encoded by a DNA sequence designated SAR-55. SAR-55 also encodes					XX						
XX		HEV ORF-2 and ORF-3 proteins. A host organism transformed or transfected					XX						
XX		with a recombinant expression vector containing the SAR-55 nucleic acid					XX						
XX		can be used to produce the HEV proteins, especially ORF-2 protein. The					XX						

CC HEV can be used for the diagnosis of infection and diseases caused by
 CC HEV, and for monitoring the progression of such disease. Such methods are
 CC also useful for monitoring the efficacy of therapeutic agents during the
 CC course of treatment of HEV infection and disease in a mammal. The
 CC antibodies can be used for detection or for passive immunisation of
 CC mammals.

XX
 SQ Sequence 1693 AA;

Query Match 82.5%; Score 7439; DB 19; Length 1693;
 Best Local Similarity 81.6%; Pred. No. 0;
 Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

QY 1 PGTTTAEQALAAASALANAVVPELSRQTELLINMOQROLVFREVLMNHQIR 60
 DB 10 pglftleagaalaanalanavvripfshqjellimqprqlvtrpewmnpidr 69
 QY 61 VINNELEYOCRARAGCLEYGAHPRSINDPNVLRHCFELRPVGRDVQRMYSAPTRGPAAN 120
 DB 70 vlnneleyocraragcleiygahprsinndpnvhrchfirlpgrdvqrvytlaptrgpaan 129
 QY 121 CRSRALRGLEPPADRYTCDFGFSRCAPAEETGVALYSLHDLMPADVAEAMARHGXTRLYAA 180
 DB 130 crsaralrgleppadrytcdfgfsrgcapaetgvalyslhdlmpadvaeamftrhgxtlryaa 189
 QY 181 LHPPEVLPBGTYHTSYLLIHSDGRAVVTYEGDTSAGCNHNVSTILRAIRTKTVGDH 240
 DB 190 lhppevlppgtyhtsylvllhdsgravvtyegdtsagcnhvnstilsrairtktygdh 249
 QY 241 PLVIEVERAIGCHFVLLTNAPEPSPMPYVPYPRSTEVYRSIFGPGSGSLTFPSACSTR 300
 DB 250 plvieriraigchfvlllnaapepmpypvprstevyrslfpgpgsltlfptscstr 309
 QY 301 STRHANPVHIMBRLMFGATLDDQAFCCSRMLMTYLRGISTKVTGALVANEGMNASEDAL 360
 DB 310 strhanpvhimbrlmfgatlddqafccsrmlmtylrgistkvltgvalvanegmnasedal 369
 QY 361 TAAITAAVLTICHOFLRQALSKGMRRLGVEHAOKFRTPLYSMTLEKSGROVTPGROQ 420
 DB 370 tavitaaylvltichoflrqaliskgmrllgvehaokftrtplysmtlekesgrdvtpgrq 429
 QY 421 FYAOCRRMLTSAGFHLDPRLVLPDESVPCRCRFLKRYAKGFCCEFMRLAGECCFLEPAA 480
 DB 430 fygacrrmltsagfhlprlvlpdesvpcrcrflkryakgfccefmrlagcccflepaa 489
 QY 481 GLVGDGHNDNEATGESVDPAPBAHLDVSGTVAVHGOLEALYRALNVPODIARASRLT 540
 DB 490 glvgdghndneatgesvdpapbaahldvsgtvavhgolealylralnvpo디아arasrlt 549
 QY 541 ANVELVASDPRLCEKRTVLGKTRFTVVOGAHLEANGPQGVYVSPDASRQSMAGSHST 600
 DB 550 anvelvasdprlcekrvtlvgktrftvvoGAHleangpqgvyspdasrqsmagshst 609
 QY 601 YELTPAGLOVRITSSNGLDCTATPFGCAPSAAGAEVAAPFCASALYRNRFQRSLRGTLW 660
 DB 610 yaasaglevryaaagldhnavfapgyvsrpsagervtafsalylrfnregrslrgtlw 669
 QY 661 LHPBGLLGTPPPSPGHIMESANPFCGEGTLTYRTWS-----TSGSSDSFSPPE 709
 DB 670 lhpbgllgtpppspghimesanpfcgegtltyrtws-----tsgssdsfsppe 724
 QY 710 AAAPMAAATPGLRHSTPRPSDIWVLRPPSEEFVDAAPV-PRAPDA-GLPGVVLTPPP 767
 DB 725 psipstiaap-----tp-----aaipppadpsptlsapargpap 761
 QY 768 PPEVHAKPSIP-PPSRNRRLITYPDGAKYVAGSLFESDCDWLVANANPGRPGGGLCHAF 826
 DB 762 gatarapaltthgathrillfypdgskvfagslfestctwvnaasvdlirpvgglchaf 821
 QY 827 YORPEPATYPTETIMREGLAAYLVLRPIIHAAVPDYRVEONDKRLAAVRETCSRRTGA 886
 DB 822 yorpespatyptetimreglaaylvlrpiihavapdyrlehnokrlaaayretcsrlga 881

QY 887 APYLLSGITQVAVVSLSPDAMEBNHRPGDELXYLTERPAAWFEANKRAQVLTITTEDTART 946
 DB 882 apyllsgitqvavvslspdamebnhrpgdelxylterpaaWfeankraqvltitTEDTart 941
 QY 947 ANLALEIDATFEGRACAGCTISPGIVHYQFTAGVPESGSGSRISIOGDVVVVVPTRELR 1006
 DB 942 anlaleidatfegracagctispgivhyqftagvpesgsgsristsiogdvvvvvptrelr 1001
 QY 1007 NSWRKGEAATPPTTAARVTIGRRVVIDEAPSRLPHLLLMQPASSVHLGDPNQIPAI 1066
 DB 1002 nswrkgEaatppttaarvtigrvviDeapsrlphlllmqpassvhlgdpnqipai 1061
 QY 1067 DEFHAGIVPAIRPELAFTSMWXYTHRCPRADVCCLINGAYKICQTTSGVLSRLFWNEPATG 1126
 DB 1062 dfhagivpaIrpeLaftsmwxythrcpradvcclingaykicqttsGvlsrlfwnepatg 1121
 QY 1127 OKLVXTQAAKAPNGAIVTHEAGATFETTTIATADARGLIOSRAHAIVALTRETEKC 1186
 DB 1122 oklvxtQaaKAPNGaivtheagatfetttiAtadarGLiosrahaivaltretekC 1181
 QY 1187 VILDAPGLAEVGTSDVIYNNFFLAGGEVGHRRSVIIPRGNDQNLGTLOAPPSCQISA 1246
 DB 1182 vildapglAEvgtsdviYnnfFlaggeVghrrsviiprgndqnlgtloappscqisa 1241
 QY 1247 YHGLAEELGHRAPVAAVILPCELEOGLLMYMQEELTVSSVYVFEITDVIHCRMAAPSQ 1306
 DB 1242 yhglaeelghrapvAAvIlpceleogllmymqeeltvssvyvfeitdviHcrmaapSq 1301
 QY 1307 RKAVALSTLVGRYGRFTKLYFAAHSADVRESLARIPTIGPVQATTCETELYEAMVEKGD 1366
 DB 1302 rkavalstlvgrYgrftklyfaahsadvreslaripTigpvqatTCetelyeamvekGd 1361
 QY 1367 GSAVLELDLCNRDVSRTTFQKXCNKFTTGETIAHGKVGCGISAMSKTFCALGPMWFRAT 1426
 DB 1362 gsavleldcnsrdvsrttfqkcnkfttgetiAhgkvGcgisamsktfcalgpmwfrat 1421
 QY 1427 EKELIALPPNIFRYGAVSESVFAAASGASGCMVENDSEFSDSQNNNSGLCEVYME 1486
 DB 1422 ekelialppnifryGavseSVfaaasGASGcmvendsefSDsqnnnsGlcevyMe 1481
 QY 1487 ECGMPQMLIRLYHIVASAMTILQAPKESLKGFMFKHSGEPSTLMTYNNNAIIAHGCEFR 1546
 DB 1482 ecgmpqmlirlyhivASamtilQapkeslkgfmfkHsgEPstlmtYnnnaIIahgcefr 1541
 QY 1547 DFRVAAFKGDSVYVLCSDYRQSRNAALLAGCGLKLVKDYRPIGLYAGVVAAPGLSTLPD 1606
 DB 1542 dfrvaaFKgdsyVylcsdyRqsrnaALLagcglkLVkdyrPiGLyagvvaapglstlpd 1601
 QY 1607 VVRFAGRLSEKMGKGPBERAEQRLAVCDPLRGLTVAQVCVDSVSVSGVGLVHNLI 1666
 DB 1602 vvrfagrlsekmgkGPberaeQrlavcdPLrGLtvaQvcvdsVsvsgvGLvhnli 1661
 QY 1667 GMLQTTADGKAMHETIKPVLDTNSIIOEVE 1698
 DB 1662 emlgavadvGkahfteesvKpvlDltnslIctve 1693

RESULT 9
 AAM76368
 ID AAM76368 standard; Protein; 1693 AA.
 AC AAM76368;
 XX
 DT 03-DEC-1998 (first entry)
 XX
 DE Hepatitis E virus hollow particle protein #1.
 XX
 KW hollow particle protein; virus; antibody; detection; immunoassay;
 XX infection.
 OS Hepatitis virus.
 XX

Key Location/Qualifiers
1..1693
/note="Partial sequence"

JP10234383-A.
08-SEP-1998.
28-FEB-1997; 97JP-0062445.
28-FEB-1997; 97JP-0062445.
(DENK-) DENKA SEIKEN KK.
(KOKU-) KOKURITSU YOKO EISEI KENKYUSHO.
WPI: 1998-535037/46.
N-PSDB: AAV61687.
Hepatitis E virus hollow particle poly:peptide(s) and nucleic acids
encoding it - useful for more accurate detection of HEV in samples,
using immuno-assays and nucleic acid hybridisation
Claim 10; Page 17-24; 29pp; Japanese.
This sequence represents a Hepatitis E viral hollow particle protein.
This polypeptides can be used to raise antibodies to detect HEV
infection in samples, e.g. by immuno-assay based techniques, and the
nucleic acid can be used for the same in nucleic acid hybridisation
assays. The polypeptides and nucleic acids allow more accurate
detection of HEV than previously possible.

Sequence 1693 AA;

Query Match 82.5%; Score 7437; DB 19; Length 1693;
Best Local Similarity 80.9%; Pred. No. 0;
Matches 191; Conservative 114; Mismatches 158; Indels 56; Gaps 5;

1 PGITTAIEBAALAAANSALANAVVPELRFVOTELINLMQPROLVEPEVIMNPIOR 60
10 paltalegaalaanaanaavvrflshqlellnlmqrqlvtrpevfmbhpr 69
61 VHNLEEOYCARAGKCEVGAHPSINDPNVLRHCFLRPGRDVQWYSATRGPAAN 120
70 vhnleeycaragrcleigahpsindpnvnrhcfllrpgrdqgrtqtrtptgpaan 129
121 CRRSALRGLPADRYTCDFGSRCAFAETGVALYSLHDLMPADVAEAMARXGTRLYAA 180
130 crrsalrglpadvrtcyldgfsgrcfpaetgialyslhdmppsdaeamfrgmtrlyaa 189
181 LHLPEVLPGCTYHTTSTVLLIHGDGRAVVTGEGDPSAGYNHDVSLRAMITRTKIVGDH 240
190 lhlpevllppgcytrtaasylldhgrtvtvlyegdtlsagynhdvsnllswitlckvqgh 249
241 PLVIERVRAIGCHFVLLTLAEPSPMPYVPYRSTEVVRSIFGPGSSPLFSPASCTK 300
250 plviertraigchfvllltaapepsmpyvpypstevvrsifpggspslfpscscsk 309
301 STEFAVPAHIVWRMLFEGATLDDQAFCCSRMLTYLRGISTYKTVGALVANEGMNASDEAL 360
310 stfahvpaahivwrmlfegatladdqafccsrmltylrgistykvtvgtlvanegmasedal 369
361 TFXITAAVLTICHOVRLFRQATISGKMRIRVGHAKFTIRLWSLFEKSGRXYTIGRORO 420
370 tavitaaavltichovrlfrqatiskmrirvghakffirlyswlfeksgrtyligrora 429
421 FTAOCRMLWSAGFHLDPRLVLFDESVPCRTFLKRVAKGFCFCFMRMLQOECTCELEPAE 480
430 ftagocrmlwsagfhlprlvlfdesvpcrtaflkalksfccfmrmlqgectcflgpaee 489
481 GLVGHGHDNEAYEGSEVDPAEPAHLGVTAHGHOLEALYRALNVPQDIAARASRLT 540
490 glvgghgdnayegsevdpaepahlgvtaahgholealylralnvpqdiavaragrlt 549

541 ATVELVASPDRLECRCTVLGNKTRFTTVYDGAHLNANGPEQYVLSFDAQSGKAGSHSLT 600
550 atvsvsqdgridcettlgnktrfttsfvdgavlneangperylsfdasqstmaagpslpt 609
601 YELTPAGLOVRISSNGLDCTATFPPGAPSAAPGEVAACSLYRXNRTQHSITGGIM 660
610 yaasaglevryvaagldhravfapyspsrpsagvetalsalryfrneaqhstlgnlhw 669
661 LHPGGLGIFPPSPGHIMESANPCGECTLYTRTWS-TSGSSDPSF-----PAAAPA 714
670 fhpegllglfapfsgphwesanpcgestlytrtwsexdavsparpdlglmsepslps 729
715 MAATPGLPHSPVSDIWLPPPSSEFOYDAAPVPAPDAGLPGPVYLTPPPPVHKKP 774
730 raatpctl-----aatlp-----lapdpppppsap 754
775 STPPP-----SNRRLLTYTPGKAVYACSLFESQCDMLVNASNGCHRP 819
755 aldepasgatagapaltbqtahrlllftypgsksvfagslfesactvlnasnvdhprg 814
820 GGLCHAFYORPEEAPYTFEIMREGIAAVTLPRPIIHAVADYVEQMPKRLAEAYREPT 879
815 gglchafyorppeaptytfemregiaavtlprpiihavadyveqmpkrlaeaayrec 874
880 CSRGTAAVPLIGSGIYQVPVLSFSDAMERNHRPDELYTPPAAWFEANKPAPVLTIT 939
875 csrgtlaavpllgsgilyvpvlsfisdamerhrpdelyltpaaawfeankpaprcltli 934
940 TEDIARTANALEIDAATEVGRACAGCTISGIVMHOFTAGVPGSGKSRSIQGDVDVYV 999
935 tedvartanalaidatevgracagctisgivmhoftagvpgsgksrsitqadavdvv 994
1000 VPTRELRNSMRGRFAAFTPHAAAVTIGRRVIDEAPSLPHLLLMHORSASHLLGD 1059
995 vptrelrnsmrgrfaaftphaaavtigrvrideapslphlllmhqrtaevhllgld 1054
1060 PQOIPAIDEFHAGLVPATIRPELATPSMWVTHRCPADVCELRGAVPKIOTSRVLSIF 1119
1055 pqoipaidfahaglvpatirpdlaptsmwvthrcpadvceilrgavpmtqtsrvlsif 1114
1120 WNEPAIGOKLYVTOAKANAPGATVHEAGATFETTTIATADARGLIOSSRAHAIVAL 1179
1115 wnepaigoklyvtgakaangpavtvehaagetyelttiatadargliqssrahaival 1174
1180 TRHTEKCVILDPBGLIREGISDVYVNNFTLAGGEVGMKRPVIRGNNDOLGTLQAFR 1239
1175 trhteksvildasgillrevgsdaiwnfflaagelgqirpsvlprrgnpdandvtlaafr 1234
1240 PSCQISAVYHQLAEELGHRPAPAAVLPCCPELEOGLIYMPOLATVSDSVLPELTDIVHC 1299
1235 pscqisafhqlaeelghrpavavlpccpelvoglilympeltdacdsvvtelcldvnc 1294
1300 RMAAPSORKAVLSTLVGRYGRRTKLYEAAHSDVRESLARFIPTIGVQATCELYELVEA 1359
1295 rmaapsoqrkavlstlvgryrtrtklynashsdvrsldarflpajpvgvvtlcellyelva 1354
1360 WYKGGODSNAVLELDCNNDVSRITFPQKXCKFTTGTETIANGKVGGSIMSKFTCLAF 1419
1355 mkekqdgdsnavleldcnndvsriltfqqdckfttgetiaqkvvggsimskftclaf 1414
1420 GPMFRAIEKELIALLPNPIFYGDAYEESYFAAASGAGSCWFENDSFEDSTONNFSIG 1479
1415 gpmfraiekellalppnpiyfgydaldclvsaavaakasmvfendfsefdestqnfslg 1474
1480 IECVVMEECGMPQWLIRLYHLVRSAMLLQAPKESLKGFWKKHSGEPGTLNMTVMNMAIT 1539
1475 iecvvmeeegmpqwlirlyhlrsawllqapkeslrgfwkksngsgpctlwntvmnmai 1534
1540 AHCFEFDPRVAARFKGDDSVVLCSDYROSRNAALILACGGLKTKDYPRIGIYAIVVAP 1599
1535 ahcfefdrvaarfkddsvvlcsdyrosrnaalilacgglktdyprigiyavvavp 1594


```

OY 1072 GLVPAIRBELATSMWXYTHRCPADVCELIRGAYPKIQTTSHVLSLFPMNEPAIGOKLVX 1131
DB 1067 GLVPAIRPDLPTSWHTRHPADVCELIRGAYPMIGTTSVLSLFWGEPAYGKLVF 1126
OY 1132 TOAKAANPAGATVHEAGATFETTTIATADARGLIOSSRAHAVALTRHTEKCVILDA 1191
DB 1127 tgaakpangpsvthneagaytettitlatadarglqssrahaivaltrhlekcvilda 1186
OY 1102 PGLLEVEGISDVYNNFLPAGEVGHKRPVYPRGNPNONLGTLOAFPPSCQISXHQLA 1251
DB 1187 PGLLEVEGISDAIVNFIAGGELGHPVSVIPRGNPDANDVTLAAPPSCQISAHQLA 1246
OY 1252 BELGRHPAPVAALPPCELEGGILYMPQELTSDSVVLFETLTVHNRMAAPSORKAVL 1311
DB 1247 eelgrhpyvvaavppcelegllylpqeltdcsavtfteldlvhcmmapsqrkav 1306
OY 1312 STLVRGRRKRLYEAAHSDVRESLAREPTIGVQATTCLELYELVEAMVERKQDGSAYL 1371
DB 1307 stlvgrgygrtklynashsdvreslarfipaigpyvttcelyelveamverkgdgsayl 1366
OY 1372 ELIDLCNRYSTRITFOKKCNKTTTGETTAHGVGOGISAMSTFCALGPMFRALKEKEL 1431
DB 1367 elidcnrvtstlftqkcnktttgetlahgvvgogisamstfcalgpmfralekell 1426
OY 1432 ALPPNIFYGDAYEESVFAAASGAGSCVFNDESEFSTONNFSLSLGEVAMEECGMP 1491
DB 1427 alppngvfygdafddtvsaaavaakaamvfndesefstqnnfsjleecimecgmp 1486
OY 1492 QWLIRLYHLVRSAMILLQAPKESLKGFWKKHSGEPCTLLMNTVNMNATIAHCYEFPRFVA 1551
DB 1487 qwlirlyhlvrsawillqapkeslrgfwkxhsgepctllmntvnmnavlthcydfdfgva 1546
OY 1552 AFKGSDDSVLCSDYOSNRMAALINGCGIKLVDRPRTGLVGVVAVAGTGLPVPVFA 1611
DB 1547 afkgsxslvloseytrgspgaavlaagcgiklvdrpripglvavvavpglpavvfta 1606
OY 1612 GLTSEKNMGCPGERAEORLAVCDPLRGLTNYAQVCVDVSVSVSGLVNHLGMLQT 1671
DB 1607 grltekmgpgeraeqrlavsdflrkltnvaqmcvdvsvsvsglvnllgmqla 1666
OY 1672 IADGKAHFTETTKPVLDLTNSTIIOQVE 1698
DB 1667 vadgkahftesvkpyldltnsilcrve 1693

RESULT 11
ID AAR51264 standard; Protein: 1693 AA.
XX
AC AAR51264;
XX
DT 21-OCT-1994 (first entry)
XX
DE HEV strain protein encoded by ORF-1.
XX
KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
XX antibody; detection; diagnosis; primates; stool suspension.
XX
OS Hepatitis E virus strain SAR-55.
XX
PN WO9406913-A.
XX
PD 31-MAR-1994.
XX
PF 17-SEP-1993; 93WO-US08849.
XX
PR 18-SEP-1992; 92US-0947263.
XX
PA (USSH ) US SEC DEPT HEALTH.
XX
PI Emerson SU, Purcell RH, Tsarev SA;
XX

```

```

DR WP1: 1994-118462/14.
DR N-PSDB: AA045197.
XX
PT Purified hepatitis E strain SAR-55 virus - used to develop prods.
PT for use in detection, diagnosis, vaccines and therapy of
PT hepatitis E virus infection
PS
XX Disclosure; Page 57-62; 114pp; English.
XX
CC The sequences given in AAR51264-66 are encoded by the hepatitis E virus
CC (HEV) strain SAR-55. The cDNA sequence contains three open reading
CC frames (ORFs). These proteins can be used to stimulate the production
CC of protective antibodies upon injection into a mammal that would serve
CC to protect the mammal upon challenge with wild type HEV. The proteins
CC can be used for detection and diagnosis of HEV infection. The HEV
CC SAR-55 cDNA was isolated from primates inoculated with stool
CC suspensions obtained from hepatitis E patients.
XX
SQ Sequence 1693 AA;

```

```

Query Match 82.5%; Score 7436; DB 15; Length 1693;
Best Local Similarity 81.5%; Pred. No. 0;
Matches 1396; Conservative 112; Mismatches 162; Indels 42; Gaps 7;

```

```

OY 1 PGITTAIBQAAIAAANSALANAVVVRPPLSRVOTELINLMOPROLVPRPEVLMNHPJOR 60
DB 10 pgitaiqbqaaiaaansalanavvrpplsrvgteellnlnmqprqlvtrpevfwnhprqr 69
OY 61 VIHNELEOYCRARAGRCLEVGANHPRSINDPNVHLRCLRPYGRGVOWYSAFTRGPAN 120
DB 70 vihnelelycrarsgrclevgahprsndpnvhlrcrlrpygrgvowylaftrgypaan 129
OY 121 CRRSALRGLPRADRYCYDGRSCAFEAETGALXSLHDLPADVAEAMARHGXRTLYAA 180
DB 130 crrsalrlpdradrycydgrsgcafetgalslhdmpdvaeamrfhgmrltyaa 189
OY 181 LHLPEVLLPRTVHTTSYLLIHDDRAVVTYEGDTSAGYNNHDSILRAMIRTTKIYGDH 240
DB 190 lhlpevllpprglytaasyllihdgrvrvvlyegdtsgaynnhdvsnlswlrrtkvlgdh 249
OY 241 PLVIERVRAICGFVLLLTAPERPMPYVPRSTEVYVNSIRFGSGSPSLFPACSTK 300
DB 250 plviervralchfvllylaapepsmpyvprrstevyvnslrfgsgspslfprscstk 309
OY 301 STEFHAVPVHMDRLMFGATLDDOAFCCSRIMTYLRGISYKYVAGALVANEGMNASDAL 360
DB 310 stefhavpahlwdrmlfgatlddgafccsrilmtylrghsykyvtvgalyanegmasedal 369
OY 361 TAXITAAVLTICHOBYLRTQALISKGMRLGVENHACFTITRLYSWLFEKSGRDYIPROLO 420
DB 370 taxitaaylvltichorylrtqaliskgmrlgvnhaqftrtlrlyswlfeksgrdyiprgle 429
OY 421 FYAGCRRLWSAGFHLDPRVLYFDESVPQRCRFLKLVKVAAGFCFCFMRMTGOEOTGLPEAE 480
DB 430 fyagcrrwlsagfhlprvlyfdesaapchrtalkavskfcfcfmrmtvgqetctclgpaee 489
OY 481 GLVGDHGHNDNAYEGSEVDPAEPALHDVSGTYAVGHOLEALRYALNPOTIARASRLT 540
DB 490 glvgdghndnayegevdpaepalhdvsgtyavgholealryalnpotiaraasrlt 549
OY 541 ATVELVASPDLRCHRTVLGNKTFRTTVVDGAHLEANGPEQYVLSFASRQSGAGSHSLT 600
DB 550 atvkvsygdgridcettlgnktrftrstfvdgavlengerpmlstfdaqsmaagrsfsl 609
OY 601 YELTPAGLQVRISSMGDLCTATPPGGAAPSAPRGVAAFCASLYRYNFTQHSITGGLM 660
DB 610 yaasaaglevryvaagldhravfapgvpsrpsavgtafcsalylfnreaqslstgnfw 669
OY 661 LHPGLIGIFPPSPGHIMESANPRCGGLTYTRFWS-----TSGSSPFSFSPPE 709
DB 670 fhpegllgpfapfsgphwesanprcgstlytrfswsevdavpspdpdligfts-----e 724

```



```

QY 361 TATTAATLTTCCHORYLFTQAIISKGMRLGVDBHAKFTITRLYSWLFEKSGRDYIFGRQLQ 420
Db 370 TAVTAAAYLTICHQYLTITGAIISKGMRLGVDBHAKFTITRLYSWLFEKSGRDYIFGRQLQ 429
QY 421 FYAOCRRRLSGFHLDPVILFDESVCRCRFLKKVAGKCCCFMRWJGOEOTCLLEPAE 480
Db 430 FYAGCRRLSAGFHLDPVILFDESVCRCRFLKKVAGKCCCFMRWJGOEOTCLLEPAE 489
QY 481 GLVGDHGDNEAGESEVDPAEPALHDVSGTYAVHGHOLEALRYALANYPDIAARSLRT 540
Db 490 GAVGQGDHNEAGESEVDPAEPALHDVSGTYAVHGHOLEALRYALANYPDIAARSLRT 549
QY 541 ATVELVASPDLECTVYGNKTFKTTVDGAHLEANGPEQVYLSDSARQSGAGSHSLT 600
Db 550 ATKVSGVDGRIICETLLIGNKTFKTTVDGAHLEANGPEQVYLSDSARQSGAGSHSLT 609
QY 601 YELPAGLOVRISSNGLCCTATPPPGASPAARVGAFAFCSALTYNENFOTRHSITGGLM 660
Db 610 YAASAGLEVRYAAGLHRAVFAFSPRAAGVAFCSALTYNENFOTRHSITGGLM 669
QY 661 LHPEGLIGIFPPSPGHIMESANPPCGEGLTYRTWS-TSGFSSDFSP----PEAAPA 714
Db 670 FHPEGILIGLFAFSPGHIMESANPPCGEGLTYRTWS-TSGFSSDFSP----PEAAPA 729
QY 715 MAATGCLPHSTPPVSDIWLPPPESEFOYDAAPV-PPAPDAGLPGVYLTP----- 766
Db 730 RAATPCLT-----AAPLPPPADPSPSPAPAPALPAEASGATCG 766
QY 767 PPPVHKSPISPPSRNRRLTYTPDGAKVYAGSLFESDCDMVYNSNGHRPGGGLCHAF 826
Db 767 APAATHQ-----TATNRLLIFYPDGSKYFAAGSLFESDCDMVYNSNGHRPGGGLCHAF 821
QY 827 YORPEAFYPTPEFIMREGIAATTLPRPIIHAVADRYEONPKRLAAYRQESRRGTA 886
Db 822 YGYPASFDASFTMRDGAAYTLCPRLIHAVADRYEONPKRLAAYRQESRRGTA 881
QY 887 AYPGLSGITVOYVSLSDAMERNHRPGDELYLTPAANWEPANKPAQVLTITDTART 946
Db 882 AYPGLSGITVOYVSLSDAMERNHRPGDELYLTPAANWEPANKPAQVLTITDTART 941
QY 947 ANLALEIDAAFEVGRACGCTISPGIVHYOFTAGVPGSGKRSIOGGQDVVVVPTREL 1006
Db 942 ANLALEIDAAFEVGRACGCTISPGIVHYOFTAGVPGSGKRSIOGGQDVVVVPTREL 1001
QY 1007 NSMRRCGFAATPHTAARVITGRVVIDEAPSLRPHLLLHMQRASVHLLDPPQIPAI 1066
Db 1002 NAWTRTGAFTPAARVITGRVVIDEAPSLRPHLLLHMQRASVHLLDPPQIPAI 1061
QY 1067 DFEHAGLVPAIRBELAPTSWMAVYTHRCPADVCELIRGAYPKIQTTSRVLSLFNNEPAIG 1126
Db 1062 DFEHAGLVPAIRBELAPTSWMAVYTHRCPADVCELIRGAYPKIQTTSRVLSLFNNEPAIG 1121
QY 1127 OKLYVTOAKKANPGATFVHEOGATFETETIATADARGLIQSRAHAYVALTHTKRC 1186
Db 1122 OKLYVTOAKKANPGATFVHEOGATFETETIATADARGLIQSRAHAYVALTHTKRC 1181
QY 1187 VILDPAGLRLREYISDVIVNNEFLAGGEVGHKRPVIRGNPDMLGTLOAFPPSCQISA 1246
Db 1182 VILDPAGLRLREYISDVIVNNEFLAGGEVGHKRPVIRGNPDMLGTLOAFPPSCQISA 1241
QY 1247 YHQLAEELGHRPAPVAAVPCPELEGGLLYPOELTVSDSVLVELFDYVHCMAAPSQ 1306
Db 1242 YHQLAEELGHRPAPVAAVPCPELEGGLLYPOELTVSDSVLVELFDYVHCMAAPSQ 1301
QY 1307 RKAIVSTIVGRGRRTKIYEAHSDVRESIAFTITIGPVQATTELEIYVAMERKOD 1366
Db 1302 RKAIVSTIVGRGRRTKIYEAHSDVRESIAFTITIGPVQATTELEIYVAMERKOD 1361
QY 1367 GSAVELDLNDRVRIFFQKXCNKFTTGERTIAGKVGOGISAKTFCALFGWFPRI 1426
Db 1362 GSAVELDLNDRVRIFFQKXCNKFTTGERTIAGKVGOGISAKTFCALFGWFPRI 1421
QY 1427 EKEIALLPPIFYGDAYEESVFAAASGAGSCWTFENDESEFDTONNFSLGLECYWME 1486

```

```

Db 1422 EKALLIPGVIFYGDATDVFISAAVAASAKSMWFENDISEFDSIGNISGLECAM 1481
QY 1487 ECGMPQWILIRLYHIVRSAMIILOAPKESIKGFWMKSGCEPGLTMTVMNMAIIHACYEER 1546
Db 1482 ECGMPQWILIRLYHIVRSAMIILOAPKESIKGFWMKSGCEPGLTMTVMNMAIIHACYEER 1541
QY 1547 DFRVAAEFGDSSVLCSDYROSRRNAALIIACGGLKLYDRIPIGLYAGVVAAPGLTLPD 1606
Db 1542 DFRVAAEFGDSSVLCSDYROSRRNAALIIACGGLKLYDRIPIGLYAGVVAAPGLTLPD 1601
QY 1607 VVRPAGRLSEKNMGPGPERAROLRAVCDPLRGLTNVQVCVDVSRVYGVSPGLVHNL 1666
Db 1602 VVRPAGRLSEKNMGPGPERAROLRAVCDPLRGLTNVQVCVDVSRVYGVSPGLVHNL 1661
QY 1667 GMLGTIADGKAHFTETIKPVLDTNSIIQRYE 1698
Db 1662 GMLGTIADGKAHFTETIKPVLDTNSIIQRYE 1693

```

RESULT 13

AA14615
ID AA14615 standard; Protein: 431 AA.

AC AA14615;

DT 16-JAN-1992 (first entry)

DE Encoded by ORF 1 of ET-NANB clone ET1.1 "forward" strand.

KW enterically transmitted non-A, non-B hepatitis virus; hepatitis C; HCV; E.coli strain B84; ATCC deposit number 67717; liver disease.

OS Enterically transmitted non-A, non-B hepatitis virus.

PN W09115603-A.

PD 17-OCT-1991.

PF 05-APR-1991; 91W0-US02368.

PR 05-APR-1990; 90US-0505888.

PA (GENE-) GENELABS INC. (USSH) US DEPT HEALTH & HUMAN.

PI Reyes GR, Yarborough PO, Bradley DW, Krawczynski KZ, Tam A;

DR WPI: 1991-325242/44.

PS N-PSDB; AA014410.

PT New viral proteins from non A-non-B hepatitis agent - used to

PT treat and prevent enterically-transmitted non-A non-B hepatitis

PT virus

PS Claim 3; Page 9; 117pp; English.

CC A positive clone ET1.1 was identified in a library prepared from

CC bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.

CC Both strands of ET1.1 were sequenced. One was designated the

CC "forward" strand because of statistical similarities to known

CC proteins and because the forward sequence is known to be

CC predominantly protein-encoding. Of the three possible reading frames

CC only the first (ORF 1) is uninterrupted; the other two include

CC many termination codons. See AA14616 and AA14617.

CC Sequence 431 AA;

Query Match 22.8%; Score 2057; DB 12; Length 431;

Best Local Similarity 88.4%; Pred. No. 2.8e-159;

Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

```

OY 1257 RPAPVAAVLPCEPEBOGLYMPQELTVSDSVLFELTDIVHCRMAAPSORKAVLSTLVG 1316
  |||||||
DB 1 rpyvaavlppeleqgillylpqeltcdsvvtfeltdivhcrmapsqrkavilstlyg 60
OY 1317 RYGRRTKLYEAHSDVRESIAREPTIGPVQATTCELYELVEAMVEKGDGSAVLELDLC 1376
  |||||||
DB 61 rygrtklynaashdvreslaripalipgyvtlcelyleveamvekgdgsavleldlc 120
OY 1377 NRDVSRITFFOKXCNKFTTGETIAGHKVGGISAWSKTFCALFGPWFRAIEKETIALLPP 1436
  |||||||
DB 121 nrdvsritffqkcnkfttgetiahgkvvggisawsktfcalfgpfraieketiallpp 180
OY 1437 NIFYGDAYEESVFAAIVSAGSCMFENDESEEDSTQNNFSLGLECVMEECMPQWLIR 1496
  |||||||
DB 181 gvfygdaiddvtsaavaakasmvfendfsetdsqnnfslglecimeecmpqwlir 240
OY 1497 LYHLVSAWILQAPKESLGFPMKHSSEPTLLMNTVMNNAITAHCEYERDFEVAFAFKGD 1556
  |||||||
DB 241 lyhlrsawilqapkeslrgfpmkhsgepgtllmwtvmnavlthcydfdfvfaafkgd 300
OY 1557 DSVILCSDYRQSRNAALAGCGLKIKVDYRPIGLYAGVVAPGLGTPDVVFEAGRLSE 1616
  |||||||
DB 301 dsivlcseyrqsppaavliagcglkikvdfrpigliyagvvapglgaldpvrfaagrlse 360
OY 1617 KMWGPPEERAEOURLAVCDPLRGLTNVAQVDSRVYGVSPGLVHNLIGMLQTIADGK 1676
  |||||||
DB 361 kmwgppeeraeqrlavsdflrkltnvaqmcvdsrvygvspglvhnligmlqavadgk 420
OY 1677 AHFTETIKPVL 1687
  |||||||
DB 421 ahftesvkpvl 431

```

```

RESULT 14
AAM80195
ID AAM80195 standard; Protein; 431 AA.
XX
AC AAM80195;
XX
DT 23-DEC-1998 (first entry)
XX
DE Protein encoded by the 1.33 kb DNA EcoRI insert ET1.1.
XX
KM Enterically transmitted non A non B hepatitis virus; ET-NANB;
KM Hepatitis E virus; HEV; Burma HEV isolate; vaccine;
KM diagnostic probe.
XX
OS Non A non B Hepatitis virus.
XX
PN US5824649-A.
XX
PD 20-OCT-1998.
XX
PF 07-JUN-1995; 95US-0475807.
XX
PR 25-JUL-1994; 94US-0278823.
PR 17-JUN-1988; 88US-0208997.
PR 11-APR-1989; 89US-0336772.
PR 16-JUN-1989; 89US-0367486.
PR 13-OCT-1989; 89US-0420921.
PR 05-JUL-1990; 90US-0505888.
PR 07-JUN-1995; 95US-0475807.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A;
PI Yarbough PO;
XX
DR WPI; 1998-58259/49.
DR N-PSDB; AAV66320.
XX
PT Hepatitis E virus proteins - useful for diagnosis or vaccine

```

```

PT production the virus
XX
PS Disclosure; Columns 45-48; 47pp; English.
XX
CC The present sequence is encoded by the 1.33 kb DNA EcoRI insert (forward
CC sequence) which is homologous to enterically transmitted non A non B
CC hepatitis virus (ET-NANB) (hepatitis E virus (HEV)) genome. The
CC specification describes an isolated protein which is specifically
CC immunoreactive with antibodies present in individuals infected with HEV
CC and encoded by a sequence contained in an open reading frame (ORF) of
CC an HEV genome. The genome has a sequence that is more than 70%
CC identical to the ORF1 sequence from Burma HEV isolate (AAV66321).
CC The protein is used as a vaccine and a diagnostic probe for ET-NANB.
XX
SQ Sequence 431 AA:

Query Match 22.8%; Score 2057; DB 19; Length 431;
Best Local Similarity 88.4%; Pred. No. 2.8e-159;
Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

OY 1257 RPAPVAAVLPCEPEBOGLYMPQELTVSDSVLFELTDIVHCRMAAPSORKAVLSTLVG 1316
  |||||||
DB 1 rpyvaavlppeleqgillylpqeltcdsvvtfeltdivhcrmapsqrkavilstlyg 60
OY 1317 RYGRRTKLYEAHSDVRESIAREPTIGPVQATTCELYELVEAMVEKGDGSAVLELDLC 1376
  |||||||
DB 61 rygrtklynaashdvreslaripalipgyvtlcelyleveamvekgdgsavleldlc 120
OY 1377 NRDVSRITFFOKXCNKFTTGETIAGHKVGGISAWSKTFCALFGPWFRAIEKETIALLPP 1436
  |||||||
DB 121 nrdvsritffqkcnkfttgetiahgkvvggisawsktfcalfgpfraieketiallpp 180
OY 1437 NIFYGDAYEESVFAAIVSAGSCMFENDESEEDSTQNNFSLGLECVMEECMPQWLIR 1496
  |||||||
DB 181 gvfygdaiddvtsaavaakasmvfendfsetdsqnnfslglecimeecmpqwlir 240
OY 1497 LYHLVSAWILQAPKESLGFPMKHSSEPTLLMNTVMNNAITAHCEYERDFEVAFAFKGD 1556
  |||||||
DB 241 lyhlrsawilqapkeslrgfpmkhsgepgtllmwtvmnavlthcydfdfvfaafkgd 300
OY 1557 DSVILCSDYRQSRNAALAGCGLKIKVDYRPIGLYAGVVAPGLGTPDVVFEAGRLSE 1616
  |||||||
DB 301 dsivlcseyrqsppaavliagcglkikvdfrpigliyagvvapglgaldpvrfaagrlse 360
OY 1617 KMWGPPEERAEOURLAVCDPLRGLTNVAQVDSRVYGVSPGLVHNLIGMLQTIADGK 1676
  |||||||
DB 361 kmwgppeeraeqrlavsdflrkltnvaqmcvdsrvygvspglvhnligmlqavadgk 420
OY 1677 AHFTETIKPVL 1687
  |||||||
DB 421 ahftesvkpvl 431

RESULT 15
AAM71208
ID AAM71208 standard; Protein; 431 AA.
XX
AC AAM71208;
XX
DT 30-OCT-1998 (first entry)
XX
DE Protein encoded by 1.33 kb EcoRI insert of clone ET1.1 of ET-NANB.
XX
KM Enterically transmitted nonA/nonB hepatitis virus; identification;
KM HEV; ET-NANB; detection; vaccine.
XX
OS Hepatitis virus.
XX
PN US5789559-A.
XX
PD 04-AUG-1998.
XX

```

Search completed: March 7, 2002, 14:03:47
Job time: 110 sec

PF 25-JUN-1994; 94US-0279823.
XX
XX 05-APR-1991; 91US-0681078.
PR 17-JUN-1988; 88US-0208997.
PR 11-APR-1989; 89US-0336672.
PR 19-JUN-1989; 89US-0367486.
PR 13-OCT-1989; 89US-0420921.
PR 05-APR-1990; 90US-0505888.
PR 25-JUL-1994; 94US-0279823.
XX

PA (GENE-) GENELABS TECHNOLOGIES INC.

PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A;
PI Yarbough PO;

DR WPI: 1998-446186/38.
DR N-PSDB; AAV54728.

XX
XX Hepatitis E virus DNA - useful for e.g. virus detection and viral
PT protein production

XX
XX Example 3; Columns 43-46; 45pp; English.

CC The present sequence is encoded by the 1.33 kb EcoRI insert of clone
CC ET1.1 of an enterically transmitted nona/nonb viral hepatitis agent
CC (ET-NANB). The sequence in plasmid pTZKFI(ET1.1) carried in E. coli
CC strain B4 is deposited under ATCC 67717. The nucleic acid sequence
CC may be used for identifying and sequencing the entire viral agent
CC (also referred to as HEV), detecting ET-NANB in infected samples,
CC e.g. by specific amplification of virus-derived DNA sequences
CC and for producing recombinant viral proteins for use in vaccines.

XX
SQ Sequence. 431 AA;

Query Match 22.8%; Score 2057; DB 19; Length 431;
Best Local Similarity 88.4%; Pred. No. 2.8e-159;
Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

OY 1257 RPAPVAAVLPQPELEQGLYMPQELTSDVSLVPELTDIVHCRMAAPSQRAVLSTLVG 1316
DB 1 rpvpaavlpqpeleqgllympqeltcdsvvlteltdivhcrmaapspqkavastlvvg 60
OY 1317 RYGRRTKLYEAHSDRESLARIPTIGVQATTCELYELVEMVEKGGDSAVLELDLC 1376
DB 61 rygrrtklynashsdvrsdrlarfpagvyttccelyelveamvekgdgsavleldlc 120
OY 1377 NBDVSRITFEQKXCNKFTTGETIAGKVGQGISAMSKTFCALFGPWRRIEKEILALLPP 1436
DB 121 nbdvsritffeqkcnkfttgetiabgkvqgisawsktfcalfgpwrriaekailallpp 180
OY 1437 NIFYGDATYESFAAASGASCMFENDFSEFDSQNNFSLGLECVNMEEGCPQWLIR 1496
DB 181 nifygdafddvysaavaakasmwyendfsefdsqnnfslglecavmeecmpqwlir 240
OY 1497 IYHLVRSAMILOAPKESLGFPMKHSSEPRTLMMNTVMNMAIIAHCEPRDFRVAFAKGD 1556
DB 241 yhlvrsamiolapkeslgtfpmkhssepRTLmmntvmnmahceprdfvrvaafkgd 300
OY 1557 DSVLCSIDYROSARNAALTAGGLKLKVDYRPIGLYAGVVAAGLGLTDPVVRFAGRUSE 1616
DB 301 dsvlcsidyrsgaavllagcglklkvdrpigliyagvvaaglgaldpvrfragrlse 360
OY 1617 KMWGPGPERAEQLACDFLGLTNVAQCVDVVSRYGVSPGLVHNIGMLQTTADGK 1676
DB 361 kmwpgperaeqlrlavsdflrkltnvaqmcvdvvsrygvspglvhnligmlqtadgk 420
OY 1677 AHTEETIKPVL 1687
DB 421 ahatesvxpvl 431

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:01:57 ; Search time 19.44 Seconds
(without alignments)
1965.565 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013

Sequence: 1 PGITTAIEQAALAAANSALA.....FTETIKPYLDITNSIQKVE 1698

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued Patents AA:*
2: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/5C_COMB.pep:*
7: /cgn2_6/ptodata/2/1aa/5D_COMB.pep:*
8: /cgn2_6/ptodata/2/1aa/5E_COMB.pep:*
9: /cgn2_6/ptodata/2/1aa/5F_COMB.pep:*
10: /cgn2_6/ptodata/2/1aa/5G_COMB.pep:*
11: /cgn2_6/ptodata/2/1aa/5H_COMB.pep:*
12: /cgn2_6/ptodata/2/1aa/5I_COMB.pep:*
13: /cgn2_6/ptodata/2/1aa/5J_COMB.pep:*
14: /cgn2_6/ptodata/2/1aa/5K_COMB.pep:*
15: /cgn2_6/ptodata/2/1aa/5L_COMB.pep:*
16: /cgn2_6/ptodata/2/1aa/5M_COMB.pep:*
17: /cgn2_6/ptodata/2/1aa/5N_COMB.pep:*
18: /cgn2_6/ptodata/2/1aa/5O_COMB.pep:*
19: /cgn2_6/ptodata/2/1aa/5P_COMB.pep:*
20: /cgn2_6/ptodata/2/1aa/5Q_COMB.pep:*
21: /cgn2_6/ptodata/2/1aa/5R_COMB.pep:*
22: /cgn2_6/ptodata/2/1aa/5S_COMB.pep:*
23: /cgn2_6/ptodata/2/1aa/5T_COMB.pep:*
24: /cgn2_6/ptodata/2/1aa/5U_COMB.pep:*
25: /cgn2_6/ptodata/2/1aa/5V_COMB.pep:*
26: /cgn2_6/ptodata/2/1aa/5W_COMB.pep:*
27: /cgn2_6/ptodata/2/1aa/5X_COMB.pep:*
28: /cgn2_6/ptodata/2/1aa/5Y_COMB.pep:*
29: /cgn2_6/ptodata/2/1aa/5Z_COMB.pep:*
30: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
31: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
32: /cgn2_6/ptodata/2/1aa/5C_COMB.pep:*
33: /cgn2_6/ptodata/2/1aa/5D_COMB.pep:*
34: /cgn2_6/ptodata/2/1aa/5E_COMB.pep:*
35: /cgn2_6/ptodata/2/1aa/5F_COMB.pep:*
36: /cgn2_6/ptodata/2/1aa/5G_COMB.pep:*
37: /cgn2_6/ptodata/2/1aa/5H_COMB.pep:*
38: /cgn2_6/ptodata/2/1aa/5I_COMB.pep:*
39: /cgn2_6/ptodata/2/1aa/5J_COMB.pep:*
40: /cgn2_6/ptodata/2/1aa/5K_COMB.pep:*
41: /cgn2_6/ptodata/2/1aa/5L_COMB.pep:*
42: /cgn2_6/ptodata/2/1aa/5M_COMB.pep:*
43: /cgn2_6/ptodata/2/1aa/5N_COMB.pep:*
44: /cgn2_6/ptodata/2/1aa/5O_COMB.pep:*
45: /cgn2_6/ptodata/2/1aa/5P_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7448	82.6	1693	US-08-478-507-7	Sequence 7, Appl
2	7448	82.6	1693	US-09-128-275A-7	Sequence 7, Appl
3	7439	82.5	1693	US-08-840-316-1	Sequence 1, Appl
4	7439	82.5	1693	US-08-809-523-1	Sequence 1, Appl
5	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
6	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
7	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
8	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
9	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
10	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
11	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
12	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
13	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
14	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
15	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
16	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
17	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
18	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
19	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
20	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
21	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
22	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
23	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
24	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
25	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
26	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl
27	7439	82.5	1693	US-08-471-971-1	Sequence 1, Appl

28	158.5	1.8	673	4	US-09-196-387-8	Sequence 8, Appl
29	158.5	1.8	949	4	US-09-196-387-10	Sequence 10, Appl
30	158.5	1.8	1327	4	US-09-196-387-2	Sequence 2, Appl
31	158.5	1.7	1185	4	US-09-041-886-23	Sequence 23, Appl
32	152.5	1.7	459	4	US-09-080-983-5	Sequence 5, Appl
33	148.5	1.6	3567	2	US-07-642-734C-4	Sequence 4, Appl
34	148.5	1.6	3567	2	US-08-439-009A-4	Sequence 4, Appl
35	145	1.6	1463	1	US-08-157-005-3	Sequence 3, Appl
36	145	1.6	1463	4	US-08-747-863-3	Sequence 3, Appl
37	145	1.6	3724	2	US-08-804-227C-10	Sequence 10, Appl
38	145	1.6	3724	2	US-08-804-198-4	Sequence 4, Appl
39	143.5	1.6	1315	3	US-08-899-595-3	Sequence 3, Appl
40	140.5	1.6	1255	2	US-09-080-897-4	Sequence 4, Appl
41	140.5	1.6	1255	3	US-08-899-595-1	Sequence 1, Appl
42	140.5	1.6	1255	4	US-09-323-735-4	Sequence 4, Appl
43	140	1.6	2152	4	US-09-036-987A-3	Sequence 3, Appl
44	140	1.6	2152	4	US-09-370-700-3	Sequence 3, Appl
45	139	1.5	902	1	US-08-396-479B-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-478-507-7
Sequence 7, Application US/08478507
Patent No. 6120988
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarbrough, Patricia O
APPLICANT: Bradley, Daniel W
APPLICANT: Kracyszynski, Krzysztof Z
APPLICANT: Tam, Albert
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,507
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988

ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183, 22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-507-7

Query Match 82.6%; Score 7448; DB 3; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY 1 PGTTAIEQALAAANSAANAVVPRFLSRVOTELLINLMOPROLVERPEVLMNHPIOR 60
DB 10 PGTTAIEQALAAANSAANAVVPRFLSHOOIELINLMOPROLVERPEVLMNHPIOR 69
QY 61 VIHNELOYCARAGRCLEVGAPRSINDPNVLRHCFRLPYGRDQVRYSAFTRGPAAN 120
DB 70 VIHNELEYCARAGRCLEIGHAPRSINDPNVLRHCFRLPYGRDQVRYSAFTRGPAAN 129
QY 121 CRRSALRGPRADRYTCFGFSKCAFAAETGVALYSLHLMRADVAEAMRHGXTLYAA 180
DB 130 CRRSALRGPRADRYTCFGFSKCAFAAETGVALYSLHLMRADVAEAMRHGXTLYAA 189
QY 181 LHLPEVLLPGYHTTSTYLLIHDGRAVYVEGDSAGYNIDVSLIRAMIRTKTVGH 240
DB 190 LHLPEVLLPGYHTTSTYLLIHDGRAVYVEGDSAGYNIDVSLIRAMIRTKTVGH 249
QY 241 PLVIERVRAIGHFVLLTAABEPSMPVYPRSTEVYRSIFGPGSPSLFPPSACSTR 300
DB 250 PLVIERVRAIGHFVLLTAABEPSMPVYPRSTEVYRSIFGPGSPSLFPPSACSTR 309
QY 301 STEFHAVPIIMWRMLFGATLDDQAFCCSRMTYLRGISIKYKTVGALVANEGMASEDL 360
DB 310 STEFHAVPIIMWRMLFGATLDDQAFCCSRMTYLRGISIKYKTVGALVANEGMASEDL 369
QY 361 TAVITAAAYLTICHOYRLTQALSKGRRLGVEHAOKETIRLXSWLEKSGROYIPROLO 420
DB 370 TAVITAAAYLTICHOYRLTQALSKGRRLGVEHAOKETIRLXSWLEKSGROYIPROLO 429
QY 421 FYAGCRRLWSAGFHLDPVLFDESVPCRCRTFLKKVAGKFCCEMRMLQOECTCLEPAE 480
DB 430 FYAGCRRLWSAGFHLDPVLFDESVPCRCRTFLKKVAGKFCCEMRMLQOECTCLEPAE 489
QY 481 GLVGHGHNDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVQDIAAARSKLT 540
DB 490 GLVGHGHNDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVQDIAAARSKLT 549
QY 541 ATTELVASDRLCERCTVGNKFTFTTVYDGAHLEANGPQOYLYLSPDASQSGAGSHSLT 600
DB 550 ATTELVASDRLCERCTVGNKFTFTTVYDGAHLEANGPQOYLYLSPDASQSGAGSHSLT 609
QY 601 YELTPAGLQVRISSNGLDCTATFPFGGAPSAAPGEVAACALYRYNRFTQHSLSLGLAM 660
DB 610 YELTPAGLQVRISSNGLDCTATFPFGGAPSAAPGEVAACALYRYNRFTQHSLSLGLAM 669
QY 661 LHPBELLGIFPPSPGHIMESANPCGEGTLTYRTWS-TSGSSSDSP-----PEAADA 714
DB 670 LHPBELLGIFPPSPGHIMESANPCGEGTLTYRTWS-TSGSSSDSP-----PEAADA 729
QY 715 MAATGRLHSTPPVSDIWLPPRSEFOVDADAPV-PPAPDPAGLP-GPVYLLPPPPPPVH 772
DB 730 RAATPTL-----AAPLPAPADPPSPAPAPLAPASGATAG 766
QY 773 KRISIP-PPSRNRLLTYVPDGAKEYAGSLFESDCDMLVNASNPGRPGGLCHAFYQRP 831

DB 767 APAITHQTAHRRRLFTYDPGSKVFAGSLFESTCTMLVNASNVDBHPPGGGLCHAFYQRP 826
QY 832 EAFYPIEFIMREGLAATLTLPRIITAVADRYEONPKLEAYETCSRGRTAAYPLL 891
DB 827 ASFDASFWMRDGAAYATLTPRPIIAHVAADYRLHNPRLEAAAYETCSRLGTAAYPLL 886
QY 892 GSGIYQVPSLSPDAMERNHRPDELYLTPEDANMEANKPAPVLTJEDTARTANLAL 951
DB 887 GTGIIYQVPSLSPDAMERNHRPDELYLTPEDANMEANKPAPVLTJEDTARTANLAL 946
QY 952 EIDAAATEVRACAGCTISPGIYHYOPTAGVPSGSKRSIQOGDVVYVYPTRELNSMR 1011
DB 947 EIDAAATEVRACAGCTISPGIYHYOPTAGVPSGSKRSIQOGDVVYVYPTRELNSMR 1006
QY 1012 RGFPAATPHTAARVYTGRRVYIDEAPSLPHLLLMORASVHLLGDPNOIPADIEHA 1071
DB 1007 RGFPAATPHTAARVYTGRRVYIDEAPSLPHLLLMORASVHLLGDPNOIPADIEHA 1066
QY 1072 GLVPAIRPELAPTSMMXVTHRCPADVCYLIRGAYPKIOTTSRYLSLFNNRPAIGOKLY 1131
DB 1067 GLVPAIRPELAPTSMMXVTHRCPADVCYLIRGAYPKIOTTSRYLSLFNNRPAIGOKLY 1126
QY 1132 TOAKANPGATVHBAOGATFETTTIATADARGLIOSSRAHAIVALTHTREKCYILDA 1191
DB 1127 TOAKANPGATVHBAOGATFETTTIATADARGLIOSSRAHAIVALTHTREKCYILDA 1186
QY 1192 PGLIREVIGSIDYVNNFFLAGEVGCHRSVYPRGPNQNLGLOAFPPSCQISATHQLA 1251
DB 1187 PGLIREVIGSIDYVNNFFLAGEVGCHRSVYPRGPNQNLGLOAFPPSCQISATHQLA 1246
QY 1252 EELGHRPAPVAVLPCCPELEGLLYMPOELVYSDSVLFEELTDIYHCHMAAPSOXKAYL 1311
DB 1247 EELGHRPAPVAVLPCCPELEGLLYMPOELVYSDSVLFEELTDIYHCHMAAPSOXKAYL 1306
QY 1307 SETLVGRRGRTKLYNASHSDVDRSLARFIPAIGPVQVTCETELVEAMVEKQGDGSAVL 1366
DB 1372 ELDLCNRDVSRTTFQKXNKFTTGETTAHGVGQGISMSKTFCLFEPWRAIRKELL 1431
QY 1367 ELDLCNRDVSRTTFQKXNKFTTGETTAHGVGQGISMSKTFCLFEPWRAIRKELL 1426
QY 1432 ALLPPIFYGDAYEESVFAAASGACSWFENDSEFPSTONNFSLGLECVMECEGMP 1491
DB 1427 ALLPPIFYGDAYEESVFAAASGACSWFENDSEFPSTONNFSLGLECVMECEGMP 1486
QY 1492 QWLIRLYHLVRSAMILLQAPKESLKGFWKHSGEPGTLIMNTVWMAITIAHCEFRDPRVA 1551
DB 1487 QWLIRLYHLVRSAMILLQAPKESLKGFWKHSGEPGTLIMNTVWMAITIAHCEFRDPRVA 1546
QY 1552 AFKGDSDVYLCDYRQSRNAALLIAGCGLKLVDYRPIGLYAGVYVAPRLGLTLPDVYRA 1611
DB 1547 AFKGDSDVYLCDYRQSRNAALLIAGCGLKLVDYRPIGLYAGVYVAPRLGLTLPDVYRA 1606
QY 1612 GRLSEKMGCPGERAEOLALAVCDFLRLGTNTVAOVCDVAVSRVYGVSPGLVYHNLJGMLDT 1671
DB 1607 GRLSEKMGCPGERAEOLALAVCDFLRLGTNTVAOVCDVAVSRVYGVSPGLVYHNLJGMLDT 1666
QY 1672 IADGKAHFTETIKPVLDLTNSIIORVE 1698
DB 1667 IADGKAHFTETIKPVLDLTNSIIORVE 1693

RESULT 2
US-09-128-275A-7
Sequence 7, Application us/09128275A
Patent No. 6229005
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarbrough, Patrice O
APPLICANT: Bradecy, Daniel W
APPLICANT: Krzewczynski, Krzysztof Z

APPLICANT: Tam, Albert
APPLICANT: Fry, Kirk E
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
TITLE OF INVENTION: No. 6229005-A/No. 6229005-B Hepatitis Viral Agent
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/128,275A
FILING DATE: 03-AUG-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Petibory, Joanne R
REGISTRATION NUMBER: 42,995
REFERENCE/DOCKET NUMBER: 4600-0183.24
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-128-275A-7

Query Match 82.6%; Score 7448; DB 4; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1396; Conservative 110; Mismatches .169; Indels 32; Gaps 6;

QY 1 PGTTTAEQAALAAANSALANAVVPRFLSRVQTEILLINIMOPRQVLFREVIMNHPIOR 60
DB 10 PGTTTAEQAALAAANSALANAVVPRFLSHQOIEILLINIMOPRQVLFREVIMNHPIOR 69
QY 61 VINHELEQYRARARGLLEVGAHRSINDNPNVYHRCFLPRVGDVORWTSAPTRGAAN 120
DB 70 VINHELEQYRARARGLLEVGAHRSINDNPNVYHRCFLPRVGDVORWTSAPTRGAAN 129
QY 121 CRSSALGLPADRTYCFDFESRCFAAETGVALYSLHDLMPADYADAMARHGXTRLYAA 180
DB 130 CRSSALGLPADRTYCFDFESRCFAAETGVALYSLHDLMPADYADAMARHGXTRLYAA 189

QY 181 LHLPEVLLPPGYHTHTSYLLIHDGRVAVTYEDGTSGAGNHDSILRAMITRTKIVGDH 240
DB 190 LHLPEVLLPPGYHTHTSYLLIHDGRVAVTYEDGTSGAGNHDSILRAMITRTKIVGDH 249
QY 241 PLVIERVRAIGCHEVLLTLAAPESPMPVYVPPSTEVYVRSJFGPGSGSPLEPSACSTK 300
DB 250 PLVIERVRAIGCHEVLLTLAAPESPMPVYVPPSTEVYVRSJFGPGSGSPLEPSACSTK 309
QY 301 STEHAVVHTWDRMLMFGALLDQAFCCSRMLTYLRISIKYVYGAALVANEAGMASDAL 360
DB 310 STEHAVVHTWDRMLMFGALLDQAFCCSRMLTYLRISIKYVYGAALVANEAGMASDAL 369
QY 361 TAVTTAAVLTTCORVYLTQVTAISKMBRLVENEHOKFTRLYSVLFKSGRDYPCGOLO 420
DB 370 TAVTTAAVLTTCORVYLTQVTAISKMBRLVENEHOKFTRLYSVLFKSGRDYPCGOLO 429
QY 421 FYACRRWLTSAGFLDPRVLFVDESVPQRFTFLKVVAGRCFCCMRWLGOCCTCFLEPAE 480
DB 430 FYACRRWLTSAGFLDPRVLFVDESVPQRFTFLKVVAGRCFCCMRWLGOCCTCFLEPAE 489
QY 481 GLVGDHGDNEAYEGSEVDPAEPALHDVSGTYAVHGOLEALYRALVPODIAARSLT 540
DB 490 GAVGDQCHDNEAYEGSDVDAESAISDSISYVPGTALPLVQALDPAEIVARAGRLT 549
QY 541 ATVELVASPDLRECRVLTGKTFRTVVDGAHLEANGPEQVYLSFDSRSGMGASHLT 600
DB 550 ATVELVASPDLRECRVLTGKTFRTVVDGAHLEANGPEQVYLSFDSRSGMGASHLT 609
QY 601 YELTPALQVIRISNGDLCTATPPGAPSPASAPAEVAFSCALRYRFRTOHSLTGLW 660
DB 610 YASAPALEVRYVAAAGLDHRAVFAVSPASAPAEVAFSCALRYRFRTOHSLTGLW 669
QY 661 LHPEGLIGTFPPSPGHIMESANPFCGEGLYRTWS-TSGFSSDESP-----PEAAPA 714
DB 670 FHPEGLIGTFPPSPGHIMESANPFCGEGLYRTWS-EVAPAPDLTFMSEPSIS 729
QY 715 MAATPGLPHSTPPVSDTWLPPRSEEFQVDAAPV-PAAPDAGLP-CPVYLTTPPPPPVH 772
DB 730 RAATPPL-----AATPPAPDPSPPSAPALAEPPASGATAG 766
QY 773 KPSIP-PPSRRRLLYTPDGAKYVAGSLFESDCDWLVNANSPGHRPGGICHAIFYORF 831
DB 767 APATHTQARRRLFTYPPDSKYFAGSLFESTCTWLVNANSPGHRPGGICHAIFYORF 826
QY 832 EAFYPTFEIMEGLAAYTLPRPIIHAVADPYRVEQPKRLEAYRETCSRGAAYPL 891
DB 827 ASFPASEFVMDGAAYTLPRPIIHAVADPYRVEQPKRLEAYRETCSRGAAYPL 886
QY 892 GSGIYQVPSLSFDAMERNHRPGDELXLTPEANMFENKPAQVLTTEDTARTANLAL 951
DB 887 GTGIYQVPIGSPFAMERNHRPGDELXLTPEANMFENKPAQVLTTEDTARTANLAI 946
QY 952 EIDATEVGRACAGCTISPGIVHYOTAGVAGSGKSSIOGQDVVYVPPRELNSMRR 1011
DB 947 ELDSDIVGRACAGCRVTPGVOTOTAGVAGSGKSSITQADVDVYVPPRELNSMRR 1006
QY 1012 RGFATTPHTAARTYIGRRVVIDEAPSLPRILLLHMORASSVHLSDPNOIPAIIDEHA 1071
DB 1007 RGFATTPHTAARTYIGRRVVIDEAPSLPRILLLHMORASSVHLSDPNOIPAIIDEHA 1066
QY 1072 GLVPAIRPELAPYSWXYTHRCPADVCCLRGAVPKIQTTSRVLSLFWNEPAIGOKLYX 1131
DB 1067 GLVPAIRPELAPYSWXYTHRCPADVCCLRGAVPKIQTTSRVLSLFWNEPAIGOKLYX 1126
QY 1132 TOAKKANPGATVYHEAOGATFTETITIAADAGLIQSSRAHAIVALTRHTECVIIDA 1191
DB 1127 TOAKKANPGATVYHEAOGATFTETITIAADAGLIQSSRAHAIVALTRHTECVIIDA 1186
QY 1192 PGLIREVGISDVIVNNEFLAGGEVGHARPSVYPRGNPDONLGLQAPPPSCQISAHQOLA 1251
DB 1187 PGLIREVGISDVIVNNEFLAGGEVGHARPSVYPRGNPDONLGLQAPPPSCQISAHQOLA 1246
QY 1252 EELGHRPAPVAAVLPPECELEGLIYMPQELTYSDSVLYVELDITVHCRAAPASQKRAVL 1311

```

Db 1247 EELGHRPVPVAAVLPCPELEGLLTPQLTTCDSVYTFELDIYVCHMAAPSORAVL 1306
QY 1312 STLVGRYGRRTKLYEAHSDVRESLARFPTTGPVQATTCETLVEINVAWEKODGSAYL 1371
Db 1307 STLVGRYGRRTKLYNAHSDVRLARFPAISPVQVTTCELELVAWEKODGSAYL 1366
QY 1372 ELDLCNRDVSRTTFQKXCNKFTTGETIAHGKVGOGISAMSKTCALFGFWFAIRKEIL 1431
Db 1367 ELDLCNRDVSRTTFQKXCNKFTTGETIAHGKVGOGISAMSKTCALFGFWFAIRKEIL 1426
QY 1432 ALPPPIIFGDAAEESVFAAASGACSCWFEDESEPDSTONNFSIGLECVMEECGMP 1491
Db 1427 ALPPQVVFEGDAEDDTVFSAVAAAKASWFEDESEPDSTONNFSIGLECAIMECGMP 1486
QY 1492 QMLIRLYHIVRSAMITQAPESILKGFMKHSGEPGLTNVTVMNMAIIAHCEYEPFRVA 1551
Db 1487 QMLIRLYHIVRSAMITQAPESILKGFMKHSGEPGLTNVTVMNMAIIAHCEYEPFRVA 1546
QY 1552 AFGKDDSVVLCSDYRQSRNAALIAAGCLKLVYDRPIGLYAGVVAAPGLTLPDVVRA 1611
Db 1547 AFGKDDSVVLCSDYRQSRNAALIAAGCLKLVYDRPIGLYAGVVAAPGLTLPDVVRA 1606
QY 1612 GRISSENMGPBPRAQLRAVCDPLRLTNVAQVCVDVYSRYGVSPGLVNHILGMLT 1671
Db 1607 GRISSENMGPBPRAQLRAVCDPLRLTNVAQVCVDVYSRYGVSPGLVNHILGMLT 1666
QY 1672 IADGRAHFTETIKPVLDTNFIOTRYE 1698
Db 1667 VADGRAHFTESYKPVLDLTNFIOTRYE 1693

```

RESULT 3

US-08-840-316-1

Sequence 1, Application US/08840316

Patent No. 6054567

GENERAL INFORMATION:

APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
 Applicant: Tsarev, Sergei. A., and Robinson, Robin A.

TITLE OF INVENTION: Recombinant Proteins Of

TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their

TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/840,316

FILING DATE: 11-APR-1997

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Richard W. Bork

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4255

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1693 AMINO ACID RESIDUES

```

; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
US-08-840-316-1

```

Query Match 82.5%; Score 7439; DB 3; Length 1693;
 Best Local Similarity 81.6%; Pred. No. 0;
 Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

```

QY 1 PGITTAIEQAALAAASALANAVVPRFLSRVQTELLINLMOROLVFRREVLMNPIOR 60
Db 10 PGITTAIEQAALAAASALANAVVPRFLSRVQTELLINLMOROLVFRREVLMNPIOR 69
QY 61 VINHELEIYCRARAGCLEGAHPRISINDPNVLAHCFLEPRVDQRMVSAFTGPAAAN 120
Db 70 VINHELEIYCRARAGCLEGAHPRISINDPNVLAHCFLEPRVDQRMVSAFTGPAAAN 129
QY 121 CRRSALRGLEPPADRTYCFDGFSCAPAEETGVALYSLHDIWPAVVAEAMARHGXYRLAA 180
Db 130 CRRSALRGLEPPADRTYCFDGFSCAPAEETGVALYSLHDIWPAVVAEAMARHGXYRLAA 189
QY 181 LHLPRVLLPFGYHTTSTYLLIHDGRVAVVTEGDSAGYNHDSVILRAMIRTTKIVGDH 240
Db 190 LHLPRVLLPFGYHTTSTYLLIHDGRVAVVTEGDSAGYNHDSVILRAMIRTTKIVGDH 249
QY 241 PLVIERVRAIGCHFVLLTLAAPPSPMPVYPRSTEVYRSIFGFGSPSLSPFASCSTK 300
Db 250 PLVIERVRAIGCHFVLLTLAAPPSPMPVYPRSTEVYRSIFGFGSPSLSPFASCSTK 309
QY 301 STEFAVAVHIIWDLMLFGATLDDQAFCCSRMLMTYLRGISYKVTGVALVANEGMNASDAL 360
Db 310 STEFAVAVHIIWDLMLFGATLDDQAFCCSRMLMTYLRGISYKVTGVALVANEGMNASDAL 369
QY 361 TAITTAAYLTICQRLTQALSKGMRRLGVEHAQKFTRLYSMLTEKSGROVILPRLQ 420
Db 370 TAITTAAYLTICQRLTQALSKGMRRLGVEHAQKFTRLYSMLTEKSGROVILPRLQ 429
QY 421 FYACRRMLSAGHILDPRLVFEDESVCRCRTFLKLVAGFCFCFMMLQCEGCFLEPAE 480
Db 430 FYACRRMLSAGHILDPRLVFEDESVCRCRTFLKLVAGFCFCFMMLQCEGCFLEPAE 489
QY 481 GLVGDHNDNEAVEGSEVDPAPPAHLIDVSGTYAVHGHQLEALYALNVPDIDIAARSLT 540
Db 490 GLVGDHNDNEAVEGSEVDPAPPAHLIDVSGTYAVHGHQLEALYALNVPDIDIAARSLT 549
QY 541 ATVEIYVADRLCEKTVLGNKTRTIVVQGAHLEANGPQIVLSPASQSMKAGSHSLT 600
Db 550 ATVEIYVADRLCEKTVLGNKTRTIVVQGAHLEANGPQIVLSPASQSMKAGSHSLT 609
QY 601 YELTPAGLOVRISNGIDCTATPPPGGAPSAAPGEVAAPFCASALYRNRFTORSLTGGIA 660
Db 610 YELTPAGLOVRISNGIDCTATPPPGGAPSAAPGEVAAPFCASALYRNRFTORSLTGGIA 669
QY 661 LHPGGLIGIFPPSPGHIMESANPFCGEGTLTYRTWS-----TSFGSSDFSPP 709
Db 670 LHPGGLIGIFPPSPGHIMESANPFCGEGTLTYRTWS-----TSFGSSDFSPP 724
QY 710 AAAPMAATPGLPHSTPPVPSDIWVLPPESEFOVDAAPV-PPAPDA-GLPGVYVLTTPP 767
Db 725 PSIPSRATP-----TP-----AAPIPPAPDPSPTLSAPARGEPA 761
QY 768 PPVHKPSIP-PPSRNRRLTYPOGAKYVAGSLFSPDDMLVMAENPPHREGGICHAF 826
Db 762 GATARAIPALTTHOTARRKRLFTYPPDSKYFASLSFSTCWTWLNASVNDHRGGGICHAF 821
QY 827 YORFEAFYPTFEINREGLAAYTLTPRPIIHAVAPRYEONKRRLEAAYRETCSRRTA 886
Db 822 YORFEAFYPTFEINREGLAAYTLTPRPIIHAVAPRYEONKRRLEAAYRETCSRRTA 881
QY 887 ATPPLGSGIYQVAVSLSPDAMEKRRHPRGDELVLTPEPAAMFPAKPAQVLTITTEDYART 946
Db 882 ATPPLGSGIYQVAVSLSPDAMEKRRHPRGDELVLTPEPAAMFPAKPAQVLTITTEDYART 941

```

QY	947	ANLEIDDAEVEVRACAGCTISPGIVYHOFATGAVGSEKSKSIQGGDVVVVPTRELR	1006
Dd	942	ANLIEDSATSVDGRACAGCTVTPGVVOYFTATGAVGSEKSKSIQADVDVVVPTRELR	1001
QY	1007	NSMRRGPAFTPHTARVIGRRVVIDEAPSLPRLLLHHORASSYHLIGDPQOIAI	1066
Dd	1002	NAMRRGPAFTPHTARVIOGRRVVIDEAPSLPRLLLHHORATYHLIGDPQOIAI	1061
QY	1067	DPEHAGLVPAIRPELATSMWXYTHHCPADVCELIIGAVPKIQTSRYLRSLFENNEPAIG	1126
Dd	1062	DPEHAGLVPAIRPDLATSMWHYTHHCPADVCELIIGAIMQIOTTSRYLRSLFEMEPAVG	1121
QY	1127	QKLYXTQAAKAANDGALITVEHAGAFTEFTTTIATADAGLITOSSRAHAYALTHTKAC	1186
Dd	1122	QKLYFTQAAKAANDGASTYVHEAGATYTTETTTIATDANGLITOSSRAHAYALTHTKAC	1181
QY	1187	VILDPAGLIREVGISIDYIVNFFLAGEGVXHRPSYTPRGNDONIGTLOAFPSCQISA	1246
Dd	1182	VILDPAGLIREVGISDAIVNFFLAGELIGHORPSYTPRGNDANDVITLAAFPSCETISA	1241
QY	1247	YHOLAEELGHRPAPVAAVIIPCEBLEQGLLYMFOELTVSDSVLVELTDTIYHCRMAAPSQ	1306
Dd	1242	FHELAEBELGHRPAPVAAVIIPCEBLEQGLLYLPQELTTCDSVVFTEINDIYHCRMAAPSQ	1301
QY	1307	RKAAVLSTLVGRYGRRTKLYEAHSDVRESLARIPTIGVQATTCCELYELVEAMVERKOD	1366
Dd	1302	RKAAVLSTLVGRYGRRTKLYEASHSDVRESLARIIPAIGVQATTCCELYELVEAMVERKOD	1361
QY	1367	GSATVLEDDLCNDRVSRITTFQKKCKNFTTGETIAGHKVQOGISANSKIFCALFCGWPFAI	1426

1 MEDIUM TYPE: FLOPPY DISK
2 COMPUTER: IBM PC COMPATIBLE
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: WORDPERECT 5.1
5
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/08/809,523
8 FILING DATE: 28-MAY-1997
9
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: PCT/US95/13102
12 FILING DATE: 03-OCT-1995
13
14 PRIOR APPLICATION DATA:
15 APPLICATION NUMBER: US08/316,765
16 FILING DATE: 03-OCT-1994
17
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: 07/947,263
20 FILING DATE: 18-SEP-1992
21
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Richard W. Bork
24
25 REGISTRATION NUMBER: 36,459
26
27 REFERENCE/DOCKET NUMBER: 2026-4032054
28
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (212) 758-4800
31
32 TELEFAX: (212) 751-6849
33
34 INFORMATION FOR SEQ ID NO: 1:
35
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 1693 AMINO ACID RESIDUES
38 TYPE: AMINO ACID
39 STRANDEDNESS: UNKNOWN
40 TOPOLOGY: UNKNOWN
41
42 US-08-809-523-1

RESULT 4
US-08-809-523-1
Sequence 1, Application US/08090523
Patent No. 6207416
GENERAL INFORMATION:
APPLICANT: Tsarev, Sergei. A., Emerson,
APPLICANT: Suzanne U., Purcell, Robert H.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:

Query Match	82.5%	Score	7439	DB	4	Length	1693
Best Local Similarity	81.6%	Pred.	No. 0				
Matches	1397	Conservative	111	Mismatches	162	Indels	42
				Gaps			
Qy	1	PGITTAIEQAAIAANSALANAVVVRPELSRVQTEELINIMOPRQVLFREPVIMHDIOR	60				
Db	10	PGITTAIEQAAIAANSALANAVVVRPELSHQIEIILINMQPRQVLFREPVIMHDIOR	69				
Qy	61	VIHNLEIQCARAGRCLEVAHNRSTINDNNVNHRCFLRVGVGDVQRWISAPTRGRAN	122				
Db	70	VIHNLEIQCARSRCLEIASHRSTINDNNVNHRCFLRVAGDVQRWVTAPTRGPAN	125				
Qy	121	CRSRLRCLPRADRTYCGDFGSCRAFAETGVALYSLHMDIPAVAEAMRHGXTRLYAA	180				
Db	130	CRSRLRCLPRADRTYCGDFGSCGFRAETISALYSLHMDSPSYALAMRHGXTRLYAA	188				
Qy	181	LHLPEVLLRPGTYHTTSYLLIHGDRAVVYEGDTSAGYNHDYSILRAWIRTKIIVGDH	240				
Db	190	LHLPEVLLRPGTYRTASYLLIHGDRVVYVEGDTSGYNHDVSNLRSMWIRTKVYGDH	244				
Qy	241	PLVIERVAICGHVLLLTAAPEPSMPVYVPRSTEVYASITGPGSGSILPFSACSTK	300				
Db	250	PLVIERVAICGHVLLLTAAPEPSMPVYVPRSTEVYASITGPGSPSLPFTSCSTK	303				
Qy	301	STFHAVPVHIMDRMLFEGATDDOAFCCSRIMTYLRGISYVTVGALVANGMNASDAL	366				
Db	310	STFHAVPVHIMDRMLFEGATDDOAFCCSRIMTYLRGISYVTVGTLVANGMNASDAL	368				
Qy	361	TAXITTAAYLITCHORYLTQOAKISGMRRLGVEHAOKITRLYSLWLFKSGRDYIPGRLO	420				
Db	370	TAXITTAAYLITCHORYLTQOAKISGMRRLEREHAOKITRLYSLWLFKSGRDYIPGRLO	423				
Qy	421	FYACCRMLSGFHLDPRVIVFEDSSVCRCTELTKYVAGKCCGMRRLGECGCELPAAE	488				
Db	430	FYACCRMLSGFHLDPRVIVFEDSAPCHCTAIRKAVSKFCGCMKWLGOBCTCFLOPAE	489				
Qy	481	GLVDEGHGHDNAYEGSEVDPAEPALHDVSGGVAVHGOLELYALVFWPDIDIAARSLT	540				
Db	490	GLVDEGHGHDNAYEGSDVDPAESALSDISGYYVPRGALDPLVQADLLPAEIVARARLT	544				
Qy	541	ATVELVASPDLRCRTVLGNKTFRTTVVGDGHLFANGPRDEQVILSFDSKOSMGASHSLT	600				

QY	181	LHLPPEVLPRGYHTHTSYLLHDSGRVAYTEGDSACYNIDVSLTRMPTFTKTVGDH	240
Db	190	LHLPPEVLPRGYHTHTSYLLHDSGRVAYTEGDSACYNIDVSLTRMPTFTKTVGDH	249
QY	241	PLVIERVRAICGHEVLLTAAPEPSMPYVYPRSTREYVVRISJFGGSGPSLTFPSACSTK	300
Db	250	PLVIERVRAICGHEVLLTAAPEPSMPYVYPRSTREYVVRISJFGGSGPSLTFPSACSTK	309
QY	301	STFAVAVVHITWDRMLTFGATLDQAFCCSRLMTYLGISIKYTVGALVANESGMASDAL	360
Db	310	STFAVAVVHITWDRMLTFGATLDQAFCCSRLMTYLGISIKYTVGALVANESGMASDAL	369
QY	361	TAXTAAVLTICHOVYLRTOAISKMBRLGVEHACKFTRLSWLTPEKSGRDTIGRQLO	420
Db	370	TAVTAAVLTICHOVYLRTOAISKMBRLGVEHACKFTRLSWLTPEKSGRDTIGRQLO	429
QY	421	FYACCRWLSAGHELDRLVLEDESVPCRCRTFLKRVAKGCCFMKILQOECTCELEPAE	480
Db	430	FYACCRWLSAGHELDRLVLEDESVPCRCRTFLKRVAKGCCFMKILQOECTCELEPAE	489
QY	481	GLVGDHNDHNEAVGSDVDPAEPAHLDVSGTVAVHNOLEALYRALVNOODIAAASRLT	540
Db	490	GLVGDHNDHNEAVGSDVDPAEPAHLDVSGTVAVHNOLEALYRALVNOODIAAASRLT	549
QY	541	ATVELVASPDLRECRTVLGNKTERITTVYDGDHLEANGPEOYVLESPDASQSGAGSHSLT	600
Db	550	ATVELVASPDLRECRTVLGNKTERITTVYDGDHLEANGPEOYVLESPDASQSGAGSHSLT	609
QY	601	VELTRALQVRLISSNGLDCTATPPTPGGAPBAAGVEVAACSAALYRNRTOHNSLTGLIM	660
Db	610	YAAAPAALEGRVYAAAGDHAFAVAPGVSAPASAGEVATACSAALYRNRTOHNSLTGLIM	669
QY	661	LHPEGLLIGIPPPSPGCHIMESANPCGEGELYRTWS-----TSGSSPSFPE	709
Db	670	FHEGGLLGPAPRSPGCHIMESANPCGEGELYRTWS-----TSGSSPSFPE	724
QY	710	AAAPAMAATPGLHSTPRVSDIVLPPRSEEFQVDAAPV-PPAPDPA-GLPBPVVLTPPP	767
Db	725	PSISRRAATP-----PP-----AAAPLPPAPRPPSPPLTSPAPARGEPAP	761
QY	768	PPVHKPSIP-PPSRNRLLTYPRDCAKYYAGSLFSSCDMLVNSNPPCHRGGGGJCHAF	826
Db	762	GATRAARALITHQARAKNRRLFTTPRDSKYPAGSLFESTCWLNVASNVDRHGGGJCHAF	821
QY	827	YQREPEAFYPTFIEIMREGLAUTLPRPILIHAAVPDYRYEONPKRLDEAAYRETCSRGT	886
Db	822	YQREPEAFYPTFIEIMREGLAUTLPRPILIHAAVPDYRYEONPKRLDEAAYRETCSRGT	881
QY	887	AYPLLSGITYQVPVSLFSFDMERNRNRGDELYLTERAAMFEAKRPAQVULTTEDTART	946
Db	882	AYPLLSGITYQVPVSLFSFDMERNRNRGDELYLTERAAMFEAKRPAQVULTTEDTART	941
QY	947	ANLLEIDATAEVGRACAGCTISPGIYHVOFTGVGSGSKSRISQGDVDDVVVYPRRELR	1006
Db	942	ANLLEIDATAEVGRACAGCTISPGIYHVOFTGVGSGSKSRISQGDVDDVVVYPRRELR	1001
QY	1007	NSWRRCRPAAPTPHTAARTITGRVVIDEAPSLPRLHLLHMQARSSVHLTCDPNOIPAI	1066
Db	1002	NSWRRCRPAAPTPHTAARTITGRVVIDEAPSLPRLHLLHMQARSSVHLTCDPNOIPAI	1061
QY	1067	DEHAGLVPAIRPLPATSMWXYTHHCOPDUGCLLGAATPKTQTSRVLRSLFQWNEPAG	1122
Db	1062	DEHAGLVPAIRPLPATSMWXYTHHCOPDUGCLLGAATPKTQTSRVLRSLFQWNEPAG	1121
QY	1127	OKLVHTQAAKANAANGATVIEHAGATFETTTIATADAGLIQSSRAHAIVALTREHTEK	1186
Db	1122	OKLVHTQAAKANAANGATVIEHAGATFETTTIATADAGLIQSSRAHAIVALTREHTEK	1181
QY	1187	VILDPAGBLREVGISDVYVNNFLLAGGEVGHXHPSVITPRGNPDONIGTLOAPRPSQJISA	1246
Db	1182	VILDPAGBLREVGISDVYVNNFLLAGGEVGHXHPSVITPRGNPDONIGTLOAPRPSQJISA	1241

Qy	1247	YHAAEELGHRPPAAVAVLPCPELCSOGLLYMPOCEVTSVLAVELDIYHCRMAASQ	1306
Db	1242	FHEIAEEIGHPPAAVAVLPCPELCSOGLLYLPQELITCDSVTEELDIYHCRMAASQ	1301
Qy	1307	RKAVLSTLVGRKGYGRTILYFAAHSVDRESLAREPIIPGVQATCELELVEANWKKOD	1366
Db	1302	RKAVLSTLVGRKGYGRTILYAAHSVDRLSARIPAIIGVQYTTCLEIYELBEANWKKOD	1361
Qy	1367	GSAAVELDLCNRDYSRITFFQKKCNKFTTGETIAHGKVGQGISAMSKTFCALFGWPRAI	1426
Db	1362	GSAAVELDLCSDRSRITFFQKKCNKFTTGETIAHGKVGQGISAMSKTFCALFGWPRAI	1421
Qy	1427	EKETILLPPIIFGDAYEESVFAAASVSGSCOMPENFSEPDSTONNFSGLCEYAE	1486
Db	1422	EKALLLPPQGVFGDAFEDTIVSAVAAKAAKAMVENDSEPDSTONNFSGLCEAIME	1481
Qy	1487	ECGMPOMLIRLYHLYRSAMILQAPKESLGGFWKKHSGEPTLLMNTVMNMAIIAHCEYFR	1546
Db	1482	ECGMPOMLIRLYHLYRSAMILQAPKESLGGFWKKHSGEPTLLMNTVMNMAIYTHCYDFR	1541
Qy	1547	DFRPAARKGDSVYLCSDYLRQSRNMAALLAGCGIKLVYRIRIGLYAGVVAAPGUGLTPD	1606
Db	1542	DLYQAAKRGDSVYLCSGYQSQSGAAVLLIAGCGIKLVYRIRIGLYAGVVAAPGUGALPD	1601
Qy	1607	VYFRAGRLSEKNMGPGGERAEQLRLAVCDLRLGTLTVAAQVADVVAASRYVGVSPGLVHMLI	1666
Db	1602	VYFRAGRLTEKNMGPGGERAEQRLAVSDFRLKTLTVAAQCMADVVAASRYVGVSPGLVHMLI	1661
Qy	1667	GMLQTLADGKAHFTETKPLVLDLTNSIQAYE	1698
Db	1662	EMQAVADGKAHFTESYKPVILDTJNSILCEAE	1693

```

1 RESULT      6
2 PCT-US93-08849A-1
3 ; Sequence 1, Application PC/TUS9308849A
4 ; GENERAL INFORMATION:
5 ; APPLICANT:
6 ; TITLE OF INVENTION: Recombinant Proteins Of
7 ; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
8 ; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
9 ; NUMBER OF SEQUENCES: 98
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: MORGAN & FINNEGAN
12 ; STREET: 345 PARK AVENUE
13 ; CITY: NEW YORK
14 ; STATE: NEW YORK
15 ; COUNTRY: USA
16 ; ZIP: 10154
17 ; COMPUTER READABLE FORM:
18 ; MEDIUM TYPE: FLOPPY DISK
19 ; COMPUTER: IBM PC COMPATIBLE
20 ; OPERATING SYSTEM: PC-DOS/MS-DOS
21 ; SOFTWARE: WORDPERFECT 5.1
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: PCT/US93/08849A
24 ; FILING DATE: 17-SEP-1993
25 ; PRIOR APPLICATION DATA:
26 ; APPLICATION NUMBER: US07/947,263
27 ; FILING DATE: 18-SEP-1992
28 ; ATTORNEY/AGENT INFORMATION:
29 ; NAME: William S. Feller
30 ; REGISTRATION NUMBER: 26,728
31 ; REFERENCE/DOCKET NUMBER: 2026-4032 PCT
32 ; TELECOMMUNICATION INFORMATION:
33 ; TELEPHONE: (212) 758-4800
34 ; TELEFAX: (212) 751-6849
35 ; INFORMATION FOR SEQ ID NO: 1:
36 ; SEQUENCE CHARACTERISTICS:
37 ; LENGTH: 1693 AMINO ACID RESIDUES
38 ; TYPE: AMINO ACID
39 ; STRANDEDNESS: UNKNOWN
40 ; TOPOLOGY: UNKNOWN
41 ;
42 ;
43 ;
44 ;
45 ;
46 ;
47 ;
48 ;
49 ;
50 ;
51 ;
52 ;
53 ;
54 ;
55 ;
56 ;
57 ;
58 ;
59 ;
60 ;
61 ;
62 ;
63 ;
64 ;
65 ;
66 ;
67 ;
68 ;
69 ;
70 ;
71 ;
72 ;
73 ;
74 ;
75 ;
76 ;
77 ;
78 ;
79 ;
80 ;
81 ;
82 ;
83 ;
84 ;
85 ;
86 ;
87 ;
88 ;
89 ;
90 ;
91 ;
92 ;
93 ;
94 ;
95 ;
96 ;
97 ;
98 ;
99 ;
100 ;

```


PCT-US93-08849A-1

Query Match 82.5%; Score 7439; DB 5; Length 1693;
 Best Local Similarity 81.6%; Pred. No. 0.
 Matches 197; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

```

QY 1 PCTTAEQAALAAANSAANAVVAPFLSRVOTELILNLMOPROLVPEPEVLMNPRIOR 60
DB 10 PCTTAEQAALAAANSAANAVVAPFLSRVOTELILNLMOPROLVPEPEVLMNPRIOR 69
QY 61 VHNLEBOYCRARAGRCLEVGAPRSINDPNVLRHCELRPGRDVORNYSAPTKGPAA 120
DB 70 VHNLEBOYCRARAGRCLEVGAPRSINDPNVLRHCELRPGRDVORNYSAPTKGPAA 129
QY 121 CRRSALRGPRADRYCFCGFSGRCNPAETGIALYSLHMSPSDVAEAMRHMTRLYAA 180
DB 130 CRRSALRGPRADRYCFCGFSGRCNPAETGIALYSLHMSPSDVAEAMRHMTRLYAA 189
QY 181 LHLPEVLLPCTYHTTSTYLLIHDGDAVVTYEGDTSAGYNHDSVILRAMIRTKIVGDH 240
DB 190 LHLPEVLLPCTYHTTSTYLLIHDGDAVVTYEGDTSAGYNHDSVILRAMIRTKIVGDH 249
QY 241 PEVIERVRAIGHFVLLTLTADEPSPMPYVPRSTEVYRSIFGPGSPLEPSSACSTK 300
DB 250 PEVIERVRAIGHFVLLTLTADEPSPMPYVPRSTEVYRSIFGPGSPLEPSSACSTK 309
QY 301 STEHNAVPIIMRMLFGATLDDQAFCCSRMTYLRGISYKYTVGLVLANEGNASEDL 360
DB 310 STEHNAVPIIMRMLFGATLDDQAFCCSRMTYLRGISYKYTVGLVLANEGNASEDL 369
QY 361 TAVITRAAYLTICHOURLRTOAISKGRRLGVEHAOKFIRLYSMLTEKSGRDYIPROLO 420
DB 370 TAVITRAAYLTICHOURLRTOAISKGRRLGVEHAOKFIRLYSMLTEKSGRDYIPROLO 429
QY 421 EYACORRWLSAGPHLDPRVLVFEDESVCRCRTFLKRVAKFCCFMKLGQECTCLEPAE 480
DB 430 EYACORRWLSAGPHLDPRVLVFEDESVCRCRTFLKRVAKFCCFMKLGQECTCLEPAE 489
QY 481 GLVGDHGHNEAYEGSEVNPAPBAHLVDSGTAVHGHOLEALYRANVODILAAARSRLT 540
DB 490 GLVGDHGHNEAYEGSEVNPAPBAHLVDSGTAVHGHOLEALYRANVODILAAARSRLT 549
QY 541 ATVELVASPDLRECLRTVLGNKTFRTVVDGHLLEANGPEQOYVLSFDSQSMGASHLT 600
DB 550 ATVELVASPDLRECLRTVLGNKTFRTVVDGHLLEANGPEQOYVLSFDSQSMGASHLT 609
QY 601 YELTPAGLOVRISNGLDCTATFPPEGAPSAAPGEVAAFCALYRYNRTQRHSLTGLM 660
DB 610 YELTPAGLOVRISNGLDCTATFPPEGAPSAAPGEVAAFCALYRYNRTQRHSLTGLM 669
QY 661 LHPGGLIGFPPSPGHIMESANPECGESTLYTRTWS-----TSGFSSDSPPE 709
DB 670 LHPGGLIGFPPSPGHIMESANPECGESTLYTRTWS-----TSGFSSDSPPE 724
QY 710 AAAPMAATPGLPHSTPPVSDIMVLPPESEFOVDAAPV-PPAPDPA-GLPGFVVLTPPP 767
DB 725 PSIPSAAPR-----TP-----AAAPLPAPDPSPLLSAPAREPAP 761
QY 768 PPPVHKPSLP-PPSRNRRLITYYPDGAKYAGSLFSDCDMLVANSNPGHREGGLCHAF 826
DB 762 GATAPAPATHTOTARHRLRLLFTYPPDSKVFAGSLFESTCTWLNVANSNVDHRRGGGLCHAF 821
QY 827 YORPEAFPTPEIFIMREGIAATLTPRPIIHVAVPYRVEONPKRIEALYRETCSSRGA 886
DB 822 YORPEAFPTPEIFIMREGIAATLTPRPIIHVAVPYRVEONPKRIEALYRETCSSRGA 881
QY 887 AYPPLGSGIYQVVSLSFDAMERNHRPGDELXLTPEPAMFEANKPAQVLTITTEDTART 946
DB 882 AYPPLGSGIYQVVSLSFDAMERNHRPGDELXLTPEPAMFEANKPAQVLTITTEDTART 941
QY 947 ANLALEIDATVYGRACAGCTTSPGIVHQFTAGVPGSGSKSNSIOGQDVVVVPTREIR 1006
DB 947 ANLALEIDATVYGRACAGCTTSPGIVHQFTAGVPGSGSKSNSIOGQDVVVVPTREIR 1006

```

```

DB 942 ANLALEIDATVYGRACAGCTTSPGIVHQFTAGVPGSGSKSNSIOGQDVVVVPTREIR 1001
QY 1007 NSWRRRGFAAFTPHPTAARTYIGRRVYIDEAPSLPRLLLHMHORASSVHLGDPNOIPAI 1066
DB 1002 NAMRRRGFAAFTPHPTAARTYIGRRVYIDEAPSLPRLLLHMHORASSVHLGDPNOIPAI 1061
QY 1067 DEHAGLVPAIRBELAPTSMMVYTHRCPADVCELNGAYPKIOTTSRVLSLFWNEPAIG 1126
DB 1062 DEHAGLVPAIRBELAPTSMMVYTHRCPADVCELNGAYPKIOTTSRVLSLFWNEPAIG 1121
QY 1127 OKLVYQAANKANPGALITYHNEQAGAFTEETTITATADAGLIQSSRAHAIVALTRETEK 1186
DB 1122 OKLVYQAANKANPGALITYHNEQAGAFTEETTITATADAGLIQSSRAHAIVALTRETEK 1181
QY 1187 VILDAAGLREVGISDVIYNNFFLAGCEYGHORPSYIPGNPNQDQTLQAPSPCOISA 1246
DB 1182 VILDAAGLREVGISDVIYNNFFLAGCEYGHORPSYIPGNPNQDQTLQAPSPCOISA 1241
QY 1247 YHQLAEELGHRAPVAAPVAPCPPELBOGLLYMPOELTVSDSVLVFELTDIVHCRMAAPSO 1306
DB 1242 YHQLAEELGHRAPVAAPVAPCPPELBOGLLYMPOELTVSDSVLVFELTDIVHCRMAAPSO 1301
QY 1307 KRAVJSTLVGRIGRTKLYEAHSDVRESLAFPIPIGVQATTCCLTELVEAMVEKGD 1366
DB 1302 KRAVJSTLVGRIGRTKLYEAHSDVRESLAFPIPIGVQATTCCLTELVEAMVEKGD 1361
QY 1367 GSAVLELDLCNHDVSRITFEFOKXCNKFTGETIAGHKVCGGTSAMSKTECALGPFPRAI 1426
DB 1362 GSAVLELDLCNHDVSRITFEFOKXCNKFTGETIAGHKVCGGTSAMSKTECALGPFPRAI 1421
QY 1427 EKETIALPPIYGDAYEESVFAAASGASGCMFENDESEEDSTONNESLGLCYVME 1486
DB 1422 EKETIALPPIYGDAYEESVFAAASGASGCMFENDESEEDSTONNESLGLCYVME 1481
QY 1487 ECGMPQMLIRLYHLVANSATILQAPKESLKGPMKKSHEGECTLLMNVMMATIAHCYER 1546
DB 1482 ECGMPQMLIRLYHLVANSATILQAPKESLKGPMKKSHEGECTLLMNVMMATIAHCYER 1541
QY 1547 DEFVAAFKGDSDSVLCSDRSRHMAALJAGGLKLVDRYRPIGLTAGVVAAPGLCTLPD 1606
DB 1542 DEFVAAFKGDSDSVLCSDRSRHMAALJAGGLKLVDRYRPIGLTAGVVAAPGLCTLPD 1601
QY 1607 VYRFAGRLTEKMWGPPEBAEOLRLAVSDELRLKLTVAQMCVDVYRVYGVSPGLVHNL 1661
DB 1602 VYRFAGRLTEKMWGPPEBAEOLRLAVSDELRLKLTVAQMCVDVYRVYGVSPGLVHNL 1661
QY 1667 GMLQITADGKAFHTETIRKVDLTNSIIORVE 1698
DB 1662 GMLQITADGKAFHTETIRKVDLTNSIIORVE 1693

```

RESULT 7
 PCT-US93-08849-1
 Sequence 1, Application PC/TUS9308849
 GENERAL INFORMATION:
 APPLICANT: Tsarev, Sergei A., Emerson,
 APPLICANT: Suzanne U., Purcell, Robert H.
 TITLE OF INVENTION: Recombinant Proteins of
 TITLE OF INVENTION: A Pakistani Strain of Hepatitis E and Their
 NUMBER OF INVENTION: 98
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1

```
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US93/08849
? FILING DATE: 17-SEP-1993
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/947,263
? FILING DATE: 18-SEP-1992
? NAME:
? ATTORNEY/AGENT INFORMATION:
? NAME: Bork, Richard, W.
? REGISTRATION NUMBER: 36,459
? REFERENCE/DOCKET NUMBER: 2026-4032
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 758-4800
? TELEFAX: (212) 751-6849
? INFORMATION FOR SEQ. ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1693 amino acid residues
? TYPE: amino acid
? STRANDEDNESS: unknown
? TOPOLOGY: unknown
? PCT-US93-08849-1

Query Match      82.5%; Score 7439; DB 5; Length 1693;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

QY 1 PGTTTAEQAALAAANSALANAVVPRFLSRVQTEIILNIMOPRQVLFREVLNHPDQR 60
DB 10 PGTTTAEQAALAAANSALANAVVPRFLSHQIEIILINIMOPRQVLFREVLNHPDQR 69
QY 61 VINNELEYCARAGCLEYGAHPRSTINDPNVLRHCELRPVGRDVRQWASAPTRGPAAN 120
DB 70 VINNELEYCARAGCLEYGAHPRSTINDPNVLRHCELRPVGRDVRQWASAPTRGPAAN 129
QY 121 CRSSALRGLEPADRTYCFDFSCFAFAETGVALYSLHDLMPADVAMARHGXTRLYAA 180
DB 130 CRSSALRGLEPADRTYCFDFSCFNPAETGIALYSLHDMSPSVAAAMRHGXTRLYAA 189
QY 181 LHLPEVLEPRTYHTTSTYLLIHDGRVAVVYEGDTSAGYNHDVSILRAMIRTKIYGDH 240
DB 190 LHLPEVLEPRTYHTTSTYLLIHDGRVAVVYEGDTSAGYNHDVSILRAMIRTKIYGDH 249
QY 241 PLTIEVRATGCHFLVLLTAAPSPMPVYPRSTEVYRSTIFPGSGSLPFSASSTK 300
DB 250 PLTIEVRATGCHFLVLLTAAPSPMPVYPRSTEVYRSTIFPGSGSLPFSASSTK 309
QY 301 STFHAVPVHIDRLMLFGATLDQAFCCSRMLTYLRGISYKTVGALVANEGMNASDAL 360
DB 310 STFHAVPVHIDRLMLFGATLDQAFCCSRMLTYLRGISYKTVGALVANEGMNASDAL 369
QY 361 TAYITAAVLTICHQRYLRTQALISKMRRLGVEHAKPITRLYSWLFKSGRDYIPGROLQ 420
DB 370 TAYITAAVLTICHQRYLRTQALISKMRRLGVEHAKPITRLYSWLFKSGRDYIPGROLQ 429
QY 421 FVACQCRWLSAGFHLDPRVLPVDESVPCRCRTFLKATYAGFCCEMRLLGDECTCLEPAE 480
DB 430 FVACQCRWLSAGFHLDPRVLPVDESVPCRCRTFLKATYAGFCCEMRLLGDECTCLEPAE 489
QY 481 GLVGDGHNDHEAYEGSEVDPAEPANLDVSTYAVHGHQALRYALRVLPDIIARASRLT 540
DB 490 GLVGDGHNDHEAYEGSEVDPAEPANLDVSTYAVHGHQALRYALRVLPDIIARASRLT 549
QY 541 ATVELVASPDRLCEKRYVNGKTRTIVVDGAHLLEANGPEQYVLSFSDASROSMGASHSLT 600
DB 550 ATVELVASPDRLCEKRYVNGKTRTIVVDGAHLLEANGPEQYVLSFSDASROSMGASHSLT 609
QY 601 YELTPAGLOVRISSNGIDCATATPPPGGAPSAAGEVAAPFASALYRKRFPQRHSLTGLW 660
DB 610 YASASAGLEVRYAAGLDHRAVFAAPGVSPPRSRAGEVATFASALYRKRFPQRHSLTGLW 669
QY 661 LHPEGLLGIPTTSPGHIMESAMPFCGEGLYLRRTWS-----TSGFSSDFSPE 709
```

```
DB 670 FHEPGLGPPAPFSPGHVWESANPFCGEGLYLRRTWS-EVDAVSPAQDGLGFTS-----E 724
QY 710 AAAPAMATPGLPSTTPVSDIWLPPRSEEPQVDAAPV-PPADPA-GLRGVVLTTPP 767
DB 725 PSIPSRATP-----TP-----AAPLPADDPSTLSAPARGSPAP 761
QY 768 PPPVHKPSIP-PPSRNRLLYTPDGAAYAGSLFEESDCMLVNASNPGHPPGGGLCHAF 826
DB 762 GATARAPALITHQARHRLFTYIPDGSKVFRAGSLFESTCTWLVNASVNDHRPPGGGLCHAF 821
QY 827 YOREPEAFPTETLMREGLAAYLTLPRIIHAAVDYVQONPKRLBAARETCSRRTGA 886
DB 822 YORPASFDAASFVWRDGAAYLTLPRIIHAAVDYVQONPKRLBAARETCSRRTGA 881
QY 887 AYPILGSGIYQVPSLSFDMERNHRGDELITYLEPAANPEAKPAPQVLTITEDART 946
DB 882 AYPILGSGIYQVPSLSFDMERNHRGDELITYLEPAANPEAKPAPQVLTITEDART 941
QY 947 ANLAETDAATEVGRACAGCTISPGIVHYQFTAGVPSGSRSTIQGDVDVVPVTRBELR 1006
DB 942 ANLAETDAATEVGRACAGCTISPGIVHYQFTAGVPSGSRSTIQGDVDVVPVTRBELR 1001
QY 1007 NSWRRRGFAPFTPTAARVYIGRRVYIDEAPSLPRLHLLHMRASSVHLGDPNQIPAI 1066
DB 1002 NAMRRRGFAFTPTTARVYIGRRVYIDEAPSLPRLHLLHMRASSVHLGDPNQIPAI 1061
QY 1067 DEFHAGLVPAIRPELATSMWXYTHRCPADYCELRATYKIQTSYVLSLFENEPALG 1126
DB 1062 DEFHAGLVPAIRPELATSMWXYTHRCPADYCELRATYKIQTSYVLSLFENEPALG 1121
QY 1127 QKLVYTOAKAANGATTVHEAGATFETTTIATAARGIQRSSRAHAYALRHRKEC 1186
DB 1122 QKLVYTOAKAANGATTVHEAGATFETTTIATAARGIQRSSRAHAYALRHRKEC 1181
QY 1187 VIIDAPGLREVGISDYIVNFFLAGGEVGHRRSVYLRGNPDQNLGTAQFPSCQISA 1246
DB 1182 VIIDAPGLREVGISDYIVNFFLAGGEVGHRRSVYLRGNPDQNLGTAQFPSCQISA 1241
QY 1247 YHOLAEELGHRPAPVAAVLPCCPELEOGLLYMPOELTVSDSVLVELTDVHCMAAPSQ 1306
DB 1242 YHOLAEELGHRPAPVAAVLPCCPELEOGLLYMPOELTVSDSVLVELTDVHCMAAPSQ 1301
QY 1307 RKAVLSLTVGRYGRRTKLYEAHSDVRESLARFPTIGPVQATTCELYELAEAVVERGQD 1366
DB 1302 RKAVLSLTVGRYGRRTKLYEAHSDVRESLARFPTIGPVQATTCELYELAEAVVERGQD 1361
QY 1367 GSAVLELDLNRDYSRTTFQKXCNKFTTGETIAGHYVGGISAMSKTFCALFGMPFRAI 1426
DB 1362 GSAVLELDLNRDYSRTTFQKXCNKFTTGETIAGHYVGGISAMSKTFCALFGMPFRAI 1421
QY 1427 EKELIALLPPIFYGDAYEBSVFAAASGASGVFENDSEPDSTONNLSLGEVYME 1486
DB 1422 EKELIALLPPIFYGDAYEBSVFAAASGASGVFENDSEPDSTONNLSLGEVYME 1481
QY 1487 ECGMPQLITLYHIVRSAMLIQAPKESLKGWKHSEBPGLIMNTVWNAALIAHCEFR 1546
DB 1482 ECGMPQLITLYHIVRSAMLIQAPKESLKGWKHSEBPGLIMNTVWNAALIAHCEFR 1541
QY 1547 DFRVAAFKGDDSVYLCDYROSRAAALIAAGCGLKYVDYRPIGLYGVVVAAPGLTGLTP 1606
DB 1542 DFRVAAFKGDDSVYLCDYROSRAAALIAAGCGLKYVDYRPIGLYGVVVAAPGLTGLTP 1601
QY 1607 VVRFAGRLSEKMMGPGERAEBQRLAVCDLRLGTLTNAVAQVCVAVSVRYGVSPGLVHNL 1666
DB 1602 VVRFAGRLSEKMMGPGERAEBQRLAVCDLRLGTLTNAVAQVCVAVSVRYGVSPGLVHNL 1661
QY 1667 GMLQTDGKRAHFTETIKPVLDLTNSTIQRVE 1698
DB 1662 GMLQTDGKRAHFTETIKPVLDLTNSTIQRVE 1693

RESULT 8
```

US-08-478-507-2
; Sequence 2, Application US/08478507
; Patent No. 6120988
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R
; APPLICANT: Yarbough, Patrice O
; APPLICANT: Bradley, Daniel W
; APPLICANT: Krawczynski, Krzysztof Z
; APPLICANT: Tam, Albert
; APPLICANT: Fry, Kirk E
; TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
; TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,507
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/279,823
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/681,078
; FILING DATE: 05-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/505,888
; FILING DATE: 05-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/420,921
; FILING DATE: 13-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/367,486
; FILING DATE: 16-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/336,672
; FILING DATE: 11-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/208,997
; FILING DATE: 17-JUN-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 324-0880
; TELEFAX: (650) 324-0960
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-507-2

Query Match 22.8%; Score 2057; DB 3; Length 431;
Best Local Similarity 88.4%; Pred. No. 1.8e-172;
Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

QY 1257 RAAVAAVILPCPELEGGILYMPDELTVSDVLELTVICHCRAAPSGRAVSTLVG 1316
DB 1 RVPVAAVILPCPELEGGILYMPDELTVSDVLELTVICHCRAAPSGRAVSTLVG 60

QY 1317 RYGRRTKLYEAASDVRESLARFIPTIGVQATTCELYELVEAMVEKGDSAVLELDLC 1376
DB 61 RYGRRTKLYEAASDVRESLARFIPTIGVQATTCELYELVEAMVEKGDSAVLELDLC 120
QY 1377 NRDSRIITFPQKXCNKFTTGETIAGKVGQGISAMSKTICALFGPFRAIEKTIALLPP 1436
DB 121 NRDSRIITFPQKXCNKFTTGETIAGKVGQGISAMSKTICALFGPFRAIEKTIALLPP 180
QY 1437 NIFGDVGESEVFAAASVAGSCMFENDESEPDSTQNNFSGLECVMEEGCPOMLIR 1496
DB 181 GVFYGDVGESEVFAAASVAGSCMFENDESEPDSTQNNFSGLECVMEEGCPOMLIR 240
QY 1497 LYHLVRSAMVILAPKESLGFKKHSGEPGLTLMNTVMMAAIIAHCEFRDPRVAFFKGD 1556
DB 241 LYHLVRSAMVILAPKESLGFKKHSGEPGLTLMNTVMMAAIIAHCEFRDPRVAFFKGD 300
QY 1557 DSVILCSIDYRQSRMAALIAAGCLKIKVDYRPIGLYAGVVAAGLGLDPDVYRFAGRISE 1616
DB 301 DSVILCSIDYRQSRMAALIAAGCLKIKVDYRPIGLYAGVVAAGLGLDPDVYRFAGRISE 360
QY 1617 KMWGPPEAEQRLAVCPFLGLTVNAOVYDVYSRVYGVSGVYHNLIGMLQTIADK 1676
DB 361 KMWGPPEAEQRLAVCPFLGLTVNAOVYDVYSRVYGVSGVYHNLIGMLQTIADK 420
QY 1677 AHFTETIKPVL 1687
DB 421 AHFTETIKPVL 431

RESULT 9
US-09-128-275A-2
; Sequence 2, Application US/09128275A
; Patent No. 6229005
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R
; APPLICANT: Yarbough, Patrice O
; APPLICANT: Bradley, Daniel W
; APPLICANT: Krawczynski, Krzysztof Z
; APPLICANT: Tam, Albert
; APPLICANT: Fry, Kirk E
; TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
; TITLE OF INVENTION: No. 6229005-A/No. 6229005-B Hepatitis Viral Agent
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,275A
; FILING DATE: 03-AUG-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/279,823
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/681,078
; FILING DATE: 05-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/505,888
; FILING DATE: 05-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/420,921
; FILING DATE: 13-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/367,486
; FILING DATE: 16-JUN-1989


```

Db 443 VTKAKDV-----FGETSFWES-----LKHVLSG--CCGR 471
Qy 467 WIGQECTCLEPAEGVLGHGHNENAYEGSEVDPAPRAHLDSGYAVHGHOLEALYRAL 526
Db 472 NL-----KGIDV----- 478
Qy 527 NVPDIAARASRLTATVELVASPDRLECRVTLGKTFRTVVDGAHLLEANGPEQVYLSPD 586
Db 479 -----VFTRKYVDKXYV----- 490
Qy 587 ASBOSMGASHSLTELTPLAGLOVRISNGLDCTANFPPEGASAPGCVAAFCSLYRY 646
Db 491 -----HSL-----GDI--ICD----- 499
Qy 647 NRTORHSLTGLGLMHPGELLGFPFSPGHIMESANPCGEGTLXTFTWSTSGSSDS 706
Db 500 -----VRLSPEDV-----GF----- 509
Qy 707 PPEAAPAMAAAPGLPHSTPVSVDIWLPPSEEPQV-----DAAPVPAPADPALPG 759
Db 510 -----LPSRVPPAR-----VFHDRLEVLNAGCYNRPVPT----- 543
Qy 760 PVLVLPPLPPPPVHKPSIPPPSRNRLLYYPGAKVYAGSLERSCDWTVMNSNPGHRG 819
Db 544 -----PPVEEP-----QGFDADL--W----- 557
Qy 820 GGLCHAFYORFEAFYPTFETIMREGLAAYTLFPRPIIHAV-----PDYRVEQNKRLEAY 876
Db 558 -----HATAASLPYRA-----TLQAG-- 574
Qy 877 RETCSRGTAAAPPLGSGIYQVVSLEPDAWERNHRPGDELYTERPAWFEANKPAQPV 936
Db 575 -----LNTDVKQLKITL----- 586
Qy 937 LITTEDTARTANLEIIDATEVGRACACTISP--GIYHOFTAVPGSGKRSI----- 990
Db 587 -----ENAKT-----ID-----GILTSVPLGLEMYE--GPDSGKTGLIILAL 623
Qy 991 -OQGDVVVVVPTRELNRSMRR--GFAAFPHTAARVTIGRR-----VVIDE 1036
Db 624 EAAGGKALYAPTRERLEAMDRRIKPPSASATONVA--LAILRRATACGAPATVVIDEC 681
Qy 1037 PSLP--PHLLH--MORASSVHLGDPNOIPIADFEHAGL--VPAIRP--ELAPTSMXVTH 1091
Db 682 FMRPLVVAIVHALSPSSIVLVGDVHIGIDFQGTSAWMLYRNVVQCRRTNQRK 741
Qy 1092 RCPADV--CELIRGAYPKIQTTSRVLRSJFWNEP-----AIGOKLVYXTOAKAAN--PGA 1143
Db 742 RCPADVAVATTFQSLYPCGCTTSGCVASISHVADYRNSOACTLCTOEBKSRHGAEGAM 801
Qy 1144 TYHEAGATFETETI--ATADARGLIOSSRAHATVALTRHREKCVIIDAQ--LIREG 1199
Db 802 TYHEAGRTFASVYILHNGSTAEQKLLAKRS--HLIVGTRHNLHYIDPRTGIDIRQLN 859
Qy 1200 ISDVLYNNFELAGGEVGHRR-----SVYPRGNPDONLGTLOAFPPSCQI--SAYHOLA 1251
Db 860 HS-----AKAEVFTDIPARLETTIVKPSSEVQRN--EVATIPQSATPHGAILHLR 909
Qy 1252 BELGHRP--APVAAYLPPCEBLOGLLYMPOELTVSDV--LVFELTDIVHCMAAPSO 1306
Db 910 KNFGQPPDCGCVALAKTGEVEFGGRAKINVELAEPDAPKPHRAFG--EGVQWVKVYNASN 968
Qy 1307 RKAVALSTLVGRGRRT--KLVEAASDVR-----ESLARFIPTGPQVQATTCELIYELVEM 1360
Db 969 KHQAQTLTSLRYTKRSADPLHE--AKEDYKRMNLSDRHMDVTVEDADRAVFEFOLKE 1027
Qy 1361 VEKGODGSAVLELDCNRVSRITTFQKXCNKFTTGETIAGHGVGGISAMSKTFPALG 1420
Db 1028 TORGGVVELLEPD--DPIRIDIDFLMKTQOK--VSPKPIINTKVGCGIAHSHSLNFFVLA 1084
Qy 1421 PWFRAIEKELALLPPIFYG-----DAYESVFAAAVSGAGSCQWVENDSEFSDTONNF 1476

```

```

Db 1085 AMIRILE-ELLRTGSTRVRYSNGLPDEEAMLLKANINOVPATVFSADWTEFDTAHNT 1143
Qy 1477 SLGLECVMECEGMPOWILRYLHLYVSAMILLOAPKESLKGFWKKH-----SGEPGLTW 1530
Db 1144 SELFPALLERIGTPAAAVNLFRERCGKRTLRA-----KGLSVEVDGLDSSAANTPCR 1198
Qy 1531 NTVMNNAITAHCEYFRDFRVAAEKGDSDVYLCSDYROSRAAALIAAGCIGK--LKVDYR 1587
Db 1199 NTIFSAVMULT--FRGVFAAEKGDSD--LLCSHYLRFDASRLHMGERYKTKHLKVEVQ 1255
Qy 1588 PGLVAGVYVABGLTLPVYRFA-----GR--LSE-----KNMGP----- 1621
Db 1256 KIVPYIGLLVASQVVL--DPVRSALKIFGRCYSELISKYVEAVDITKGWSDARYHSL 1314
Qy 1622 -----GPERAEOLRLAUCDFLRG 1639
Db 1315 LCHMSACYYNVAPEASAYIITIDANVREGRC 1343

```

```

RESULT 11
US-08-093-453B-2
; Sequence 2, Application US/08093453B
; Patent No. 5439814
; GENERAL INFORMATION:
; APPLICANT: Frey, Terry K.
; APPLICANT: Dominguez, Geraldina
; APPLICANT: Wang, Chin-yeu
; TITLE OF INVENTION: Modified Infectious Rubella Virus
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jamie L. Greene, Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: United States
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/093,453B
; FILING DATE: 19 JUL 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 07/722,334
; FILING DATE: 28 JUN 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 07362-0101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404,818-3700
; TELEFAX: 404,818-3799
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEtical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Rubella virus
; STRAIN: Thelien
US-08-093-453B-2

```

```

Query Match 4.2%; Score 374.5; DB 1; Length 2205;
Best Local Similarity 19.2%; Pred. No. 3,7e-23;

```


QY 1345 -----PVQATTCELYELVEMAVE--KGODGSAYLE-----LDLCNRD 1379
Db 2116 CFRKYACNDEYMEFARKPIRITTT-----BEVTAIVARLKGPRAALFAKTHNLVPLQEP 2171
QY 1380 VSRITFFQKXCKNFTTG--ETIAHGKVGCGISMSKTFCCALFGPWFRALKEITALLPPN 1437
Db 2172 MDRFVMDKRDVAVTPGTHTHERPKV-QVIAAEPLATAYLCGIHRELVRRLTAVLLPN 2230
QY 1438 I---FYGDAVEESVFAAASGAGSCWVFENDSEPDSTONNFSLGLECVMEECGMPWL 1494
Db 2231 IHTLFMSAEFDPAITAEHFKQGD-PVLETDIASFDSODDAMALGLMLTEDLGVDPL 2289
QY 1495 IRL---YHLVRSAMILQAPKESLKGFWKKHSGEPGLLNNFYVNMALIAHCE--FRDF 1548
Db 2290 LDIECAFGEISSTHLP.TGTRFRFGAMK--SGMFLLFVNTVLTAVIASRVLEERLKS 2347
QY 1549 RVAAFKGDSDSV--LCSDYRQSRNAALIAAGGLK-----VDYRPIGLYAGVVAAPGL 1601
Db 2348 KCAAFIGDNIIGHVYVSDKEMAECAATWL--NMEVKIIDAVIGERPPYFCGGFILDV 2404
QY 1602 GTLPDVVRFAGRLSE--KNWGPGR---ERAEOLRLAVCD-----FLRGLTNVAQVCVDV 1651
Db 2405 TS--TACRYADPLKRLFKLGLPLPADEQEDERRRALLDETKAMFRVGIITDLAVAVATR 2462
QY 1652 SRVYVSPGLVHNLIGMLQTIADGKAHF 1679
Db 2463 YEVDNITPVLL-----ALRTFAQSKRAF 2485

Search completed: March 7, 2002, 14:02:59
Job time: 62 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:01:57 ; Search time 29.2 Seconds
(without alignments)
4429.602 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013

Sequence: 1 PGIITTAEOAALAAANSALA.....FTETIKPVLDLNTSIORVE 1698

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7448	82.6	1693	1 MNWVHE	genome polyprotein
2	7433	82.5	1691	1 A44212	genome polyprotein
3	1368	15.2	290	2 C48547	nonstructural prot
4	763	8.5	152	2 B38196	probable RNA-direc
5	391.5	4.3	1646	1 WMTMS2	186k protein - cuc
6	390.5	4.3	2115	2 S36480	nonstructural prot
7	374.5	4.2	2205	1 MNWVRN	non-structural poly
8	372	4.1	141	2 A38196	non-structural hyp
9	349	3.9	1844	2 S01956	hypothetical prote
10	346	3.8	1985	2 S19151	hypothetical prote
11	324	3.6	1616	1 JQ2144	183k protein - tom
12	323	3.6	1615	1 WMTW8T	180k protein - tom
13	320	3.6	1597	2 S65053	genome polyprotein
14	314	3.5	1601	2 S48699	178k protein - tob
15	313	3.5	1611	1 WMTMPV	183k protein - pep
16	311	3.5	1748	1 JQ1555	genome polyprotein
17	310	3.4	1844	1 RRPPTM	genome polyprotein
18	305.5	3.4	1776	1 RRPPTM	genome polyprotein
19	296	3.3	1615	1 WMTM18	183.3k protein - t
20	293.5	3.3	1839	1 RRPPEM	genome polyprotein
21	293	3.3	1874	1 JQ0533	genome polyprotein
22	260	2.9	1608	1 WMTGM	183k protein - tob
23	243.5	2.7	1707	2 S01865	genome polyprotein
24	243.5	2.7	1884	1 A45353	genome polyprotein
25	242.5	2.7	1885	2 JQ2183	hypothetical 216.5
26	227	2.5	1112	2 S49432	replicase 126k - o
27	226.5	2.5	1766	2 S03701	141k protein - pea
28	218	2.4	2512	1 MNWVS	nonstructural poly
29	211	2.3	2514	1 MNW82	nonstructural poly

30	207	2.3	2514	1 MNWVNZ	nonstructural poly
31	202.5	2.2	2492	1 C44213	nonstructural poly
32	202	2.2	1729	2 A49282	fusion protein 1a/
33	201.5	2.2	1707	2 S77910	hypothetical prote
34	201.5	2.2	1718	1 JQ1734	genome polyprotein
35	200	2.2	2431	1 MNWVSF	nonstructural poly
36	194	2.2	1968	1 PN0093	genome polyprotein
37	184.5	2.0	3175	1 RRPVEV	genome polyprotein
38	184	2.0	1707	2 S77908	hypothetical prote
39	183	2.0	1116	2 S26358	hypothetical prote
40	177.5	2.0	2492	1 MNWVTD	nonstructural poly
41	176	2.0	1385	2 S34230	156k protein - pla
42	174.5	1.9	2492	1 A44213	nonstructural poly
43	172.5	1.9	2493	2 S26372	nonstructural poly
44	169.5	1.9	2493	2 S72349	nonstructural poly
45	167	1.9	810	1 PZWMBR	2a protein - broad

ALIGNMENTS

RESULT 1

MNWMHE
genome polyprotein - hepatitis E virus (strain Burma)
N:contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: hepatitis E virus
C:Date: 30-Sep-1992 #sequence-revision 30-Sep-1992 #text-change 23-Jul-1999
C:Accession: A40778; A48547
R:Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, Virology 185, 120-131, 1991
A:Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length
A:Reference number: A40778; MID:92024067
A:Accession: A40778
A:Molecule type: genomic RNA
A:Residues: 1-1693 <TAM>
A:Cross-references: GB:M32400; NID:q330023; PIDN:AAA5734.1; PID:q330024
A:Title: sequence extracted from NCBI backbone (NCBI:104572, NCBI:104573)
A:Reference number: A48547; MID:92271462
A:Accession: A48547
A:Molecule type: genomic RNA
A:Residues: 967-1693 <PRY>
A:Cross-references: GB:M32400; NID:q330021; PIDN:AAA03206.1; PID:q330022
A:Note: sequence extracted from NCBI backbone (NCBI:104572, NCBI:104573)
C:Superfamily: hepatitis E virus nonstructural protein
C:Keywords: ATP; nonstructural protein; nucleotidyltransferase

Query Match	82.6%	Score 7448;	DB 1;	Length 1693;
Best Local Similarity	81.8%	Pred. No. 0;		
Matches 1396;	Conservative 110;	Mismatches 169;	Indels 32;	Gaps 6;
OY	1	PGIITTAEOAALAAANSALNAVVRPELSRQVTEILINMOPROVIFREVLNMHTOR	60	
DB	10	PGITTAEOAALAAANSALNAVVRPELSHQVTEILINMOPROVIFREVLNMHTOR	69	
OY	61	VINNELEOYCARAGRCLEYGAPHSINDPNVLRHCFRPGVDVORWYSAPTRGPAAN	120	
DB	70	VINNELELYCARASGRLEIGAPHSINDPNVYHRCFLRPGVDVORWYTAPTRGPAAN	129	
OY	121	CRSSALRGLPADRTYCEDFGFSRCAFAETGVALYSIHDMIPADVAEMARHGXTRLYAA	180	
DB	130	CRSSALRGLPADRTYCLDFGSCNFPALFTGIALYSIHDMSPDVAEMARHGXTRLYAA	189	
OY	181	LHPPEVLLPFGYHTHTSYLLIHDDGAVVYTGSDTSAGYNNHVSILRAIRTKIVGDH	240	
DB	190	LHPPEVLLPFGYHTHTSYLLIHDDGAVVYTGSDTSAGYNNHVSILRAIRTKIVGDH	249	
OY	241	PLVIERVRAIGCHFVLLITAAPESSPMYPVYPRSTFVVYRSIFGPGSGSLFPSSACSTK	300	
DB	250	PLVIERVRAIGCHFVLLITAAPESSPMYPVYPRSTFVVYRSIFGPGSGSLFPSTSCSTK	309	

```

OY 301 STEFHAVPIHIDRLMFGATLDDQAFCCSRMTYLRIGSYKYTVGALVANEGMNASEDAL 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 310 STEFHAVPIHIDRLMFGATLDDQAFCCSRMTYLRIGSYKYTVGALVANEGMNASEDAL 369
OY 361 TAXITAAVYITICHOIRLQOAISSKRRRLGVHAQKFTIRLIXSWLEKSGROYIPROQO 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 TAXITAAVYITICHOIRLQOAISSKRRRLGVHAQKFTIRLIXSWLEKSGROYIPROQO 429
OY 421 FYAOCRRMLTSAFPHIDPRVLVDESVPCRCRTEFLKVKAGFCCFMWLOECCFLEPAE 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 430 FYAOCRRMLTSAFPHIDPRVLVDESVPCRCRTEFLKVKAGFCCFMWLOECCFLEPAE 489
OY 481 GLVGDHGDNEAYEGSEVDPAEPAHLVSGTYAVHGHOLEALYRALNVODIAAARSRLT 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 490 GAVGOGHNEAYEGSEVDPAEPAHSALISDSIGSYVVPCTALQPIYQALDPAEYARAGRIL 549
OY 541 AAVELVASDRLCECRVIGNKTEPTTVYDGAHLEANGPQOYLYSPDASQSMGASHSLT 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 AAVKVSQYDGRIDCEFLGNKTFRTSEVDGAVLETNGPERNHLSEPASOSTMAAGFSLT 609
OY 601 YELTFAGLQVRISSNGLDCTATFPPGAPASAPGEVAACSAALYRNRFTORHSLTGLM 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 610 YASASAGLEVRYVAAGLDRAVAFAPGVSPRSAPGEVTACSCALYRNRQAQHSILGNLM 669
OY 661 LHPEGLLGIFFPSPGHIWESANPCGEGTLYTRTWS-TSGFSSDFSP-----PEAAPA 714
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 670 FHPEGLIGLFAFSPGHWESANPCGEGTLYTRTWS-EVADVASPARPLDGFMSSEPSIBS 729
OY 715 MAATGCLPHSTPPVSDIWLPRPSEEFQYDAPV--PRAPDPAGLP-GRVYLVLP PPPPPH 772
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 730 RAATPLT-----AAPLPAPADPPSPAPALAPASGATAG 766
OY 773 KPSIP-PPSRNRRLTYTPDCAKYVAGSLFESDCMVLVANSNPHRGPGGLCHAFYORPP 831
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 767 APAIHQARHRRRLFTYDGSKVAFAGSLFESTCTMVLVANSVNDHRPGGLCHAFYORRP 826
OY 832 EAFVTEFIMREGLAAYTLTPRIIHAAVADYRVENONPRLEAAVRETCSRRTAAVPLL 891
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 837 ASFDASAFEMRDGAAYTLTPRIIHAAVADYRVLEHNPRLEAAVRETCSRRTAAVPLL 886
OY 892 GSGIYOVNVSLEFDMERNHRPGDELYLEPAAANFEAKKPAQVLTITTEDARTANLAL 951
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 887 GGTIGYOVNPGSEFDMERNHRPGDELYLEPAAANFEAKKPAQVLTITTEDARTANLAL 946
OY 952 EIDAAATEVRACAGCTISPGIYVHOFTAGVPSSGKSRSIOGDVYVVPVRELNRSMR 1011
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 947 ELDASTDVGRACAGRVTPRGVVOYQFTACVPSSGKSRSITQADVYVVPVTELRNMR 1006
OY 1012 RGEAFTPHTAARVYTGRRVVIDEAPSLPHLLLLHMOBASSVHLLGDENQIPALDEFHA 1071
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1007 RGEAFTPHTAARVYTOGRRVVIDEAPSLPHLLLLHMOBAAVHLLGDENQIPALDEFHA 1066
OY 1072 GLVPAIRELPAPTSMWXYTHRCRADVCELIRAGVYKIQTSVNVLSLPMNEAIOGLVX 1131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1067 GLVPAIRELPAPTSMWXYTHRCRADVCELIRAGVYKIQTSVNVLSLPMNEAIOGLVX 1126
OY 1132 TOAARAPAGATVHEAOGAFTETTTIATADARGLIOSSRAHAVALTRHEKCVIIDA 1191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1127 TOAARAPAGATVHEAOGAFTETTTIATADARGLIOSSRAHAVALTRHEKCVIIDA 1186
OY 1192 PGLAREVISDYVNNFLAGGEVGHRSVTPRGNDPONTLQLOAFPSCOISAYHOLA 1251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1187 PGLAREVISDYVNNFLAGGEVGHRSVTPRGNDPONTLQLOAFPSCOISAYHOLA 1246
OY 1252 EELGHRPAPVAAVLPCEBLEGGILYMPQELTVSDSVLFEJLTDIVHCRMAAPSORKAVL 1311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1247 EELGHRPAPVAAVLPCEBLEGGILYMPQELTVSDSVLFEJLTDIVHCRMAAPSORKAVL 1306
OY 1312 STINGRVGRKRLYAAASDVRESIARFTPTIGVQATTCELXYELVEAMVEKGOGDSAVL 1371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1307 STINGRVGRKRLYAAASDVRESIARFTPTIGVQATTCELXYELVEAMVEKGOGDSAVL 1366
OY 1372 ELDLCNRDVSRLTFQKCKNFTTGETIAHGKVGOGISAMSKTFCALGPWFRALKEKEL 1431

```

```

Db 1367 ELDLCNRDVSRLTFQKCKNFTTGETIAHGKVGOGISAMSKTFCALGPWFRALKEKEL 1426
OY 1432 ALLPPIRIFGDAYEEVSFAAASGASCMVFENDSEEPSTONNEFLGLECYVMECCGP 1491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1427 ALLPQGVFYGDADFDTVFSAVAAAASAVFENDSEEPSTONNEFLGLECYVMECCGP 1486
OY 1492 QMLIRLHLVRSAMILQAKESILGFMKHSGBEPGLNNTVWNNMILIAHCEYFDFPRA 1551
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1487 QMLIRLHLVRSAMILQAKESILGFMKHSGBEPGLNNTVWNNMILIAHCEYFDFPRA 1546
OY 1552 AFGDSDSVVLCSDYROSNAALIIAGCGILKLVDPICLYAGVVAPGLGLTLDVYRRA 1611
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1547 AFGDSDSVVLCSDYROSNAALIIAGCGILKLVDPICLYAGVVAPGLGLTLDVYRRA 1606
OY 1612 GRLEKNGMGPBRRAEQLAVCDPLRGILTNNVQOVVDVSVRYGVSPLVHNLICMLDT 1671
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1607 GRLEKNGMGPBRRAEQLAVCDPLRGILTNNVQOVVDVSVRYGVSPLVHNLICMLDT 1666
OY 1672 IADGRAHFEETIKPVLDLTFNSIIORVE 1698
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1667 VADGRAHFEETIKPVLDLTFNSIIORVE 1693

```

```

RESULT 2
A44212
genome polypeptide - hepatitis E virus (strain Mexico)
N:Contains: RNA-directed RNA polymerase (PC 2.7.7.48)
C:Species: hepatitis E virus
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 19-Jan-2001
C:Accession: A44212; B48547
R:Huang, C.C.; Nguyen, D.; Fernandez, J.; Yun, K.Y.; Fry, K.E.; Bradley, D.W.; Tam, A
Virology 191, 550-558, 1992
A:title: Molecular cloning and sequencing of the Mexico isolate of hepatitis E virus
A:Reference number: A44212; MUID:93079657
A:Accession: A44212
A:Molecule type: genomic RNA
A:Residues: 1-1691 <HUA>
A:Cross-references: GB:M74506; NID:9330017; PID:AAA45730.1; PID:9330018
R:Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Piatak, M.; F
virus Genes 6, 173-185, 1992
A:title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region e
A:Reference number: A48547; MUID:92271462
A:Accession: B48547
A:Molecule type: genomic RNA
A:Residues: 965-1691 <FRY>
A:Note: Sequence extracted from NCBI backbone (NCBI:104576, NCBI:104578)
C:Superfamily: hepatitis E virus nonstructural protein
C:Keywords: ATP; GTP binding; nonstructural protein; nucleotide binding; nucleotidylt
F:973-980/Region: nucleotide-binding motif A (P-loop)
F:979/Binding site: ATP/GTP (Lys) #status predicted

```

```

Query Match 82.5%; Score 7433; DB 1: Length 1691;
Best Local Similarity 81.0%; Pred. No. 0;
Matches 1384; Conservative 121; Mismatches 166; Indels 38; Gaps 5;

OY 1 PGITTAIEQALAAANSALANAVVYRPLSRQOTELINLMOPRQOLVREPEVLMNHPOR 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 PGITTAIEQALAAANSALANAVVYRPLSRQOTELINLMOPRQOLVREPEVLMNHPOR 69
OY 61 VHNLEDOYCRARAGRCLEVGANPRISINDNPVLRHCFPRVGRVORRYSAPTGSPAN 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70 VHNLEDOYCRARAGRCLEVGANPRISINDNPVLRHCFPRVGRVORRYSAPTGSPAN 129
OY 121 CRSAALRGLPADRTYCEFDGSRCAFAAETGALSLDMLPADVAAEMARHGXTRLYAA 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 CRSAALRGLPADRTYCEFDGSRCAFAAETGALSLDMLPADVAAEMARHGXTRLYAA 189
OY 181 LHLPEVLLPGTYHTSYLLIHGDRAVVTYEGDTSAGYNHDSILRAMIKTRIVGDH 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 LHLPEVLLPGTYHTSYLLIHGDRAVVTYEGDTSAGYNHDSILRAMIKTRIVGDH 249

```

```

QY 241 PVTIERAIGCHFLLTLTAPEPSMPYVYPRSTEVYASIFPGPGSPSLFSPASCTK 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 250 PLVIERARGICGHVLLTLTAPEPSMPYVYPRSTEVYASIFPGPGSPSLFSPACAVK 309
QY 301 STEHAVVHIMDLMLFGATLDDQAFCCSRLMTYLRGISYKVTVGALVANGMNASDAL 360
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 STEHAVVHIMDLMLFGATLDDQAFCCSRLMTYLRGISYKVTVGALVANGMNASDAL 369
QY 361 TAVTAAYLTICHOBYLTQTAISKGMRLYEHAOKFTIRLYSMLEFESGDIYIPGRLO 420
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 370 TAVTAAYLTICHOBYLTQTAISKGMRLYEHAOKFTIRLYSMLEFESGDIYIPGRLO 429
QY 421 FYACCRMLSAGFHLDPRLVLFEDSVPCRCFTFKVAGKCCCMRMVLGOCCTGLEPAE 480
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 430 FYACCRMLSAGFHLDPRLVLFEDSVPCRCFTFKVAGKCCCMRMVLGOCCTGLEPAE 489
QY 481 GLVGDHNDENAYEGSEVDAPEPAHLDVSGTYAVAHQLEALYRALNYPQDIAARSLT 540
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 490 GLVGDHNDENAYEGSEVDAPEPAHLDVSGTYAVAHQLEALYRALNYPQDIAARSLT 549
QY 541 ATVELVASPDRECRIVLGNKTFRTYVDGAHLEANGBEQYVLSFDSRQSGAGSHSLT 600
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 550 ATYVVTTSGRLDQCTMIGKTFLLTFVDGARLEVNGBEQNLNLSFDSQCSMAAGPCLT 609
QY 601 YELTPAGLOYRISNGLDCTATPPGAPSAPEVAAPFCALRYNNRFTGRHSILTGLW 660
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 610 YAAVDGLEHFEHSTAGLESRYVPPGNAPTAPSEVTAFCALYRHNOSOROSVIGSLW 669
QY 661 LHPBGLIGIFPPSPGHIMESANFCEGLTYRTWTST-----SGFSSDSPPE 709
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 670 LHPBGLIGIFPPSPGHIMESANFCEGLTYRTWTSTIDTFLTVGLTSG-HIDAPHS 728
QY 710 AAPAMATGELPHSTPPVSDIWLPPSEFQYDAAPVPAPDPLGPGVLTLPPPP 769
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 729 GGPATATGAVGSDSDPDD--PLPDTGSRPSGAR-PGPNPNGVP----- 774
QY 770 PVHKPSLPPSRNRLTYTPDGAKKYAGSLFESDCDMLVNASPNRPGGLCHAFYOR 829
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 775 -----QRLHHTYPDGAKIYVGSFSESECTVLVNASNAGHRPGGLCHAFYOR 822
QY 830 FPEAFYTEREIMREGLAAYTLTPRPIIHAAVADYRVEDONPKLEAAARETSSRGTAAYP 889
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 823 YPSPFDATKFEVMDGLAAYTLTPRPIIHAAVADYRLEHNPRLEAAARETCAARGTAAYP 882
QY 890 LLSGGLYQVYVSLSEFDMERNHRPDELTYLTERPANNFEANKPAQVLTITEDTARTNL 949
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 883 LLSGGLYQVYVSLSEFDMERNHRPDELTYLTERPANNFEANKPAQVLTITEDTARTNL 942
QY 950 ALEIDATVEYGRACAGCTISPGIYHOFYTAGVPSGSKRSIQGDDVVVVVPTRELNSW 1009
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 943 ALEIDSGSEYGRACAGCTISPGIYHOFYTAGVPSGSKRSIQGDDVVVVVPTRELNSW 1002
QY 1010 RRRGFAFTPTHTAARVITGRVVVIDEAPSLPHLLLHMORASSVHLLGDNPOTPAIDFE 1069
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1003 RRRGFAFTPTHTAARVITGRVVVIDEAPSLPHLLLHMORASSVHLLGDNPOTPAIDFE 1062
QY 1070 HAGLVAIPRELAPTSWXYVTHRCPADVCELINGAYPKIOTTSVLSLFEWNEBAGOKL 1129
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1063 HTGLIPIRELPVTSWMVHTHRCPADVCELINGAYPKIOTTSVLSLFEWNEBAGOKL 1122
QY 1130 VYVQAAAPAAAPGATVHEAGATFTTTITATADARGLIOSSRAHAIVALTRHEKVCYL 1189
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1123 VYVQAAAPAAAPGATVHEAGATFTTTITATADARGLIOSSRAHAIVALTRHEKVCYL 1182
QY 1190 DAGGLLEVGISDVIVNFFLAGEVGXHRPSVIRPGNPONLGTLOAFPPSCISAYHO 1249
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1183 DAGGLLEVGISDVIVNFFLAGEVGXHRPSVIRPGNPONLGTLOAFPPSCISAYHO 1242
QY 1250 LABELGHRPAPVAAVLPCEPELEGLIYMPQDELTVSDSVLVELTIDIVHCRMAAPSQRKA 1309
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1243 LABELGHRPAPVAAVLPCEPELEGLIYMPQDELTVSDSVLVELTIDIVHCRMAAPSQRKA 1302
QY 1310 VLSTIVGRYGRRTKIVEAASDVRESLARIPTIGPVQATTCELVELVEAMVEKQDQSA 1369

```

```

|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1303 VLSTIVGRYGRRTKIVEAASDVRESLARIPTIGPVQATTCELVELVEAMVEKQDQSA 1362
QY 1370 VLEIDLGNDRVSRITFEFOKXCNKFTTGETIAHGKVGQISAMSKTCALFGWFAIREKE 1429
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1363 VLEIDLGNDRVSRITFEFOKXCNKFTTGETIAHGKVGQISAMSKTCALFGWFAIREKE 1422
QY 1430 ILALTPRIEFYGDAAVEEVEFAAASGACSVFENDESEFPSTONNFSLGLECYMEBCG 1489
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1423 ILSTLPQAVFEGDAVDOSVFSAAGAASHAVFENDESEFPSTONNFSLGLECYMEBCG 1482
QY 1490 MPQWLRLYHLVRSAMIIQAKRESILKGFWKHSGEPGLLNNTVYNNMIIAHCFEPRDR 1549
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1483 MPQWLRLYHLVRSAMIIQAKRESILKGFWKHSGEPGLLNNTVYNNMIIAHCFEPRDR 1542
QY 1550 VAAFKGDDSVVLCSDYRQSRNAALIAGCGKLKLYADYRPIGLYAGVVAAPGIGLTPDYVR 1609
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1543 VAAFKGDDSVVLCSDYRQSRNAALIAGCGKLKLYADYRPIGLYAGVVAAPGIGLTPDYVR 1602
QY 1610 FAGRLSEKNMGPGPERAQLAVCDLRLTNNAQVCVDVYSRYGVSPGLVHNLIGML 1669
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1603 FAGRLSEKNMGPGPERAQLAVCDLRLTNNAQVCVDVYSRYGVSPGLVHNLIGML 1662
QY 1670 OTIADGRAHFTETIKPVLDTNLSIIOBYE 1698
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1663 OTIADGRAHFTETIKPVLDTNLSIIOBYE 1691

RESULT 3
C48547
nonstructural protein - hepatitis E virus (strain Tashkent) (fragment)
C:Species: hepatitis E virus
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 20-Sep-1999
C:Accession: C48547
R:Fly, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Platak, M.;
Virus Genes 6, 173-185, 1992
A:Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region
A:Reference number: A48547; MUID:92271462
A:Accession: C48547
A:Molecule type: genomic RNA
A:Residues: 1-290 <FRY>
A:Cross-references: GB:110337; NID:q291457; PID:AAA4573.1; PID:q291458
A>Note: sequence extracted from NCBI backbone (NCBIN:104577, NCBIPI:104580)
C:Superfamily: hepatitis E virus nonstructural protein
C:Keywords: ATP; nonstructural protein

Query Match 15.28; Score 1368; DB 2; Length 290;
Best Local Similarity 85.98; Pred. No. 1.6e-78;
Matches 249; Conservative 21; Mismatches 20; Indels 0; Gaps 0;

QY 1344 GPVQATTCELVELVEAMVEKQDQSAVLEIDLGNDRVSRITFEFOKXCNKFTTGETIAHGK 1403
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GPVQATTCELVELVEAMVEKQDQSAVLEIDLGNDRVSRITFEFOKXCNKFTTGETIAHGK 60
QY 1404 VGQGISAMSKTCALFGWFAIREKEILALLPNIIFYGDAAEESVFAAASGACSVNFE 1463
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VGQGISAMSKTCALFGWFAIREKEILALLPNIIFYGDAAEESVFAAASGACSVNFE 120
QY 1464 NDESEFPSTONNFSLGLECYMEBCGMQWLRLYHLVRSAMIIQAKRESILKGFWKHSG 1523
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NDESEFPSTONNFSLGLECYMEBCGMQWLRLYHLVRSAMIIQAKRESILKGFWKHSG 180
QY 1524 EPGTLNNTVYNNMIIAHCFEPRFAVAAFGGDSVVLCSYRQSRNAALIAGCGKLK 1583
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 EPGTLNNTVYNNMIIAHCFEPRFAVAAFGGDSVVLCSYRQSRNAALIAGCGKLK 240
QY 1584 VDRPPIGLYAGVVAAPGIGLTPDYVRPAGRLSEKNMGPGPERAQLAV 1633
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 VGRFPIGLYAGVVAAPGIGLTPDYVRPAGRLSEKNMGPGPERAQLAV 290

RESULT 4

```



```

QY   1234 TLQAEPSPSOISATYHQLAEELGHRAPVAANVLPCEBLEGLLMPDELTVSDSVLVEFL 1293
      | :                               | : | |
Db    1134 TM-----FATVPTRQLMNSTLY----- 1151
QY   1294 TDIIVCHRMAPPSSORKAVLSITVGRTGR-----RKLTYAHSDVRESL 1336
      || : | | | | | | | | | | | | | | | | | |
Db    1152 ---VHNIIPLVPSKGTGYTDMOEFDYDRCILPGNSFVLNDFDAVTMRLDNEFNILQPCRLTL 1208
QY   1337 ARETIP-----IGPVATPCE-----LYELEVMEVKKGODGSALV-ELDICNR 1378
      | : | | | | | | | | | | | | | | | | | |
Db    1209 SNIDVPALIKNEQNLIKPIRLTACEPRPRIPELLLENVAMIKRNMTPLDAGTDITNM 1268
QY   1379 DVS-----RTTFQKKXCKNFETTGETIANGKVG----- 1405
      | : | | | | | | | | | | | | | | | | | |
Db    1269 SISIYDNFESSFVRDEVLDHDICVARASTIQSTSDMFSCOPTSAVGOLANFNFLDPAFD 1328
QY   1406 -----OGISAKSKFCALFGPMWRRAIEKETILLPPN--I 1438
      | : | | | | | | | | | | | | | | | | | |
Db    1329 TYWHMIKRQPKSRDLDTISQSEYPALOTIYHPRVVANAVEPVRKYLLTKRLSLWDSKFF 1388
QY   1439 FTGDAAEESY--FAAAVSGAGSCMFENDESFEDSTONNESLGEVCVMEECGMQWLIR 1496
      || : | : | : | : | : | : | : | : | : | : |
Db    1389 FYTRKKPEPDLOEFFSPLSSHSDYEILELDVSKYDKKSQSDHFESIEMAIWEKLGIDITLAW 1448
QY   1497 LVHIYVASNILQAPKRSIAG--FWKKHSGEGPTLLNMT-----YNNMALIAHCYERDRF 1548
      :: : | | | : | : | : | : | : | : | : | : |
Db    1449 MMSMGHKRTILDFOAGIKFTLIYYORRSSGVTTFIGFTFIACVASMLEDKCFK--- 1504
QY   1549 RVAAFEGLDDSVYLCS---YROSRNAAILLAGGKLKLVYRPDLGIYAGVVV---APGLG 1602
      | : | | | | | | | | | | | | | | | | | |
Db    1505 --ASFEGCDSLIYLPKGLEYPDIOATANLV--WNFEKLRFRKKRYGCGKTIHHANGCT 1560
QY   1603 TLPDVFVRFAGRISEKNMGCPGERAEQDLRLAVCDFLGLGTNVVA 1644
      || : | : | | | | | : | : | : | : | : |
Db    1561 YRPDPLKLSKLGKNSL-VGEYHEEFRRISLDVAHSIFNGA 1601

RESULT        6
S38480
nonstructural protein - rubella virus
C:Species: rubella virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
J:Accession: S38480
R:Gillam, S.
submitted to the EMBL Data Library, March 1993
A:Description: Nucleotide sequence of the nonstructural protein genes of rubella virus
A:Reference number: S38480
A:Accession: S38480
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-2115 <GtL>
A:Cross-references: EMBL:X72393; NID:g410507; PIDN:CAA51087.1; PID:g410508
C:Superfamily: rubella virus nonstructural polyprotein

```

	Query Match	4.3%	Score 390.5	DB 2;	Length 2115;
	Best Local Similarity	19.3%	Pred. 5.8e-167		
	Matches 452;	Conservative 205;	Mismatches 733;	Indels 947;	Gaps 105;
OY	3 ITTATGQAALAAANASALANAVVPPELSRVQTEIINIMQPRLVREVLNHHRIORI	62			
	: :	: :	: :	: :	
Db	40 VYTAQOKRAIV-----AVIPRFVPTOMQVS-----DHF--AL	69			
OY	63 HNELEOYCRARAGCLEVGAMHPRS-----INNPNVLHRCLRPGRDVORWYS-----	111			
	: :	: :	: :	: :	
Db	70 H-AISRYTRR--HWIEWG--PKAHLVLIIDSPGL--LREAARERRVALCLHRT	118			
OY	112 --APRRGPAN--CRSALKGLPP-----ADPFTYC-PFGGF	141			
	: :	: :	: :	: :	
Db	119 ARKLATALETRSEGMHADYYC--ALRKABSGPFYVHPEDVPHCGRAVAADRCLLYTPM	175			
OY	142 SRCGAETGVALYSLEHMDPADVA-----E	167			
	: :	: :	: :	: :	

Dd	176	OMELMRTIDATLTVAVDMLPVALAAHVGDMDDLGIAIHLHDHGGCCPADRCGACAGFPP	235
Oy	168	AMAHGRTRLYAALHLEPEVLLPPGYHTHTSYLLIHGDGRAVTEYEGDTSAGYNHDVSL	227
Dd	236	GTYRPTCTRITQVL---PDIAHGRGLRCGRPLMTFDCAVAEISMENVAOHGCHQ---	286
Oy	228	RAMIRTKIVGDHPIEVRVAIGCHFVLLLTAAPEPSDMPVVPYPRSTEVYRSIFPG	287
Dd	287	-----ARBAVRCTPLIRHVRSLDPS-----ANVR-----	311
Oy	288	GSPSLFBSACSTSTFPAVNVVHIDRMLMEFGATLDDQAFCCSRMLTYLGRISYKYTVAL	347
Dd	312	-LPLVHLAEVGMWMMFSLPRPVEORHLSCKYTLSPDAYYSERVKFKNALSHSITLAGN	370
Oy	348	VANGMWN---ASEDALAXITAAVYLITICHRYLRTQALISKMBRPLGVEAOKFTRILXSW	404
Dd	371	VLOGMKGTCAEDCALCAVAFR---AMOSNARLAGIKSKAKCAADSLS-----VAGW	421
Oy	405	LEFEKSGNDYIPGRLOFYA-----OCRRMLSGFHLDPRLVLPFDESVPBCRCPFLKVA	458
Dd	422	L-----DLTMDATKRFSGSVPLAEHMEEM-----EQDAVAANFD-----	455
Oy	459	GKFECCPMKWLGOECTCLPEABGLVDGHGHDNEATVEGSEVDPABEPAHLDVSTYAVHGO	518
Dd	456	-----RGPLEDGGRHLDTVQ-----PKSPPREPLATWLVHAAS	490
Oy	519	LE---ALYRLNVPOD-----IAARASLT	540
Dd	491	ADRHCAACPRCDVFRERPSAPAGRPDDDEALIPWILFAEBRRALRCREMDFEALARAADTAA	550
Oy	541	ATVELVASPDRLCECTVL-----GNKTRTIVVGGAHLEANGLEANGPOVYLIS	584
Dd	551	APAPLARPRARYP---TVLYNHPAHNHGPWILLDEPGGADALVLCRPLQPLRGPRRH---	605
Oy	585	FDASRQSMGAGSHLYELTPAQIQ---VRISNGLDCTATPPGGAPSAAGEVAAFCSA	642
Dd	606	-----YAAGAHMCA---OARGLOAFVVR-----VPPERPMWADG---ARAWAK	643
Oy	643	LYRNRTORHNSLUGJMLPBGLLG---IPPPSPGHIMESANPFCSEGLTYLRMTST	698
Dd	644	FFRCACAAOR-----LLEBPVAMHLPRYDGGVPRKLT-----ALALRTLIAQ	683
Oy	699	SGFSSDS-----PREAAPAPMAATPG-LPHSTRPVSDIWLPPP-----SEEF	741
Dd	684	OGAALALSVNDLPRGTAFENANATYAVVRAGPQLATSPRPD-----PPPRRARARSQH	739
Oy	742	QVDAAPVPRAP--DPAGLPGRVVLTPPPPPVHKPSIRPP-----SNRRLLTYTY--PD	791
Dd	740	SDAGSTPAPVPRDPP--RPQP---SPAPRPVQDP--VPPTTAEPADRARNAHELLEVUYEPS	794
Oy	792	GAKYAGASLESFD-----CDMLVUNASNGCHRGCGGLCAAF	826
Dd	795	GPPTSTADPDSIDVESYARAAGPVHILRVARDIMDPPGCKVUNANANBSLAGSGVCAI	854
Oy	827	YOREPFAFYPTFETIMREGIAA---YTLTPRP-----IIHAAVAP-----DY	863
Dd	825	FANATAA-----LAADCRRLAPCIGEAVATPCHGGGTYHIIHAAVARRPRDPA	903
Oy	864	RVEONPKRLAEARE-----TCSRKGTAAVPLIGSITY---QYVPLISPDAMERNHRPGDE	916
Dd	904	ALEGEALLERAYRSIVALAAARRMARVACPLIGAGVYGSMAESERIALAATRAEPAPER	963
Oy	917	LYL-----TEPAAWPFANRPADQVYLITTEDRT	946
Dd	964	VSLIHCIPDRATLTHASVLYGAGLAARRVSPPTPELAS--CEAGDPRGAORSASPATPR	1022
Oy	947	ANLALEIDAATEVGRACAGCTIS-----PGIV-----	973
Dd	1023	LG-----DATAPEBRCGGCGELCRLRYTRVINDRAYVWMLERBGRATSMAMRLPEVVUYGR	1077
Oy	974	-----HYQF-----TAGVPGSGSKR-----	988
Dd	1078	EHIAHTPLPLNYSLKPAEVRPRPGMGCGDMMHRCRMOGMPQVCRTPSNAAHALCRITGP	1137

```
QY 989 ---SIOGQDVVVVPTRELIN-SMRRGFAAFTPHTAARTIGR----- 1029
| : : : |
Db 1138 PAVSTRGELDENCTCMLRAAANVAQAARACGAYTSAGCKAYGRALSERTHEDEPAALS 1197
QY 1030 -----RVVIDEAPSLPHLLHMQA----- 1051
| : : : |
Db 1198 OMMSASHADASPDGTGDLPLMETVGCACSRVWVSGSEHEAPDHLVSLHRAHPGPGV 1257
QY 1052 -----SSVHL-----GDPNQI-----PAI----- 1066
| : : : |
Db 1258 VLEVARRREGNPTGHFVCAVGGPRRVSDRPHMLAVPLSGGTCATDEGLAQAYD 1317
QY 1067 DEE-----HAGLVPAPRELAAPS--WXYVT----- 1095
| : : : |
Db 1318 DLEVRLGDADAMAARALASIORPRKGPYIRVWNNMAGAKTTRILAAFTREDLYCPTN 1377
QY 1096 ---DYCELR-----GAYPKI-----QTSR 1113
| : : : |
Db 1378 ALHHEIOAKLRARDIDIKNAATYERATKPLAAYRIYIDEAFTLGGEYCAVASQTTAE 1437
QY 1114 VL-----RSLF--WNEP----- 1123
| : : : |
Db 1438 VICVGDROCCGPHYANNCTPYPDMPKPTGRSRHTMRFPCMAARLRAGLDYIEGERTGT 1497
QY 1124 -----AIGQKLVYQAAKAAANPGAITVHEAGATFETTTIATADARGL- 1167
| : : : |
Db 1498 FACNLMDGROVDLHLAFSREYV--RRLHEAGIRATVREAQMSVGTACIHVGROGTDVA 1555
QY 1168 IOSSAAHAIVALTRHEKCVIID-APGLIREVIGSDVYVNNFFLAGGEV-----GXH 1218
| : : : |
Db 1556 LALTDLAIIVSLTRASDLYLHELEDGLLRAGLS-----AFLDAGALAELEKVPAGID 1609
QY 1219 R-----PSVPRGNPNQNLGTLQAPRPSQISAVHOLAEELGHRPAPVAIVLPCREL 1271
| : : : |
Db 1610 RVAAVEQAPPLP---PADGIPEADQDVPPCPRTLEIEIFEGRAGH---PHYVDLNVYTG 1663
QY 1272 EOGGLYMPQELTSDSVVLFELTDIVHCHMAAPSORKAVLSTLVGRRTLYEAAHSD 1331
| : : : |
Db 1664 EREVAYMR-----ISRHLNKNHTEMPTERYLSAVSCGRAGE-----DGST 1707
QY 1332 VRESIARLP---TIGVQAT---CELYELVAMVEKQDGSVALEL----- 1373
| : : : |
Db 1708 LRTAAROHPRFROIPPRPVTAGVAQEWMTYLREERIDLTDTVYQMGVAARELIDRYTR 1767
QY 1374 -----DLGNRDSRIITFEQKXCKF-----TTGETIAHKGVGGCISAMSTPCA 1417
| : : : |
Db 1768 RYPEIFAGMCTAOSLSVPAFLKATLKCYDALGPRDTECHAAQGAAGLEIRAMAQEWQ 1827
QY 1418 LFGPWRAIEKEIILALPPNIFYGDAYESVFAAASGAGSCMVFENDESEFSTQNNFS 1477
| : : : |
Db 1828 VMSPIFRALOKITMALRQFLVAGHTEPEVDAMQAHYITNAIEVDTEEDMNTLAT 1887
QY 1478 -----LGECVAMECCGPMOWLIRLYHVS--AMLOAPRESIKGFKKHSGEP 1525
| : : : |
Db 1888 RDVELEISAAALGLPCA--ED-----YRALRAGSYCTLRLEJSTETCERSTGCP 1935
QY 1526 GTLLMNTYNNMNAIIAHCEFR---DERVAA--FKGDDSVLSCDVRQSNAA---ALI 1575
| : : : |
Db 1936 ATLENNTT---VAMCAMRVNPKGVWAGIFQGDVNIPLPE--GAINAALKTWTPAV 1988
QY 1576 AGCGKLKLVYD--RPIGLYAGVVAAPGLGLTDPVVFAGRLSEKMMGPPREAPQRLAV 1633
| : : : |
Db 1989 GLFGHHIYVKHSTPTSPSCG--HVGTAAGLFHDVNHQAIKVCRRP--DPDVLBEQVAL 2045
QY 1634 CDPLRGLTNVAVCVDVVSRVYGVSPGLVNHNLGMLQTTIADKAHETETIKVLDLTNSI 1693
| : : : |
Db 2046 IDRLRG-----YYAALP-----DIVAANAAYDYSAERVLAIIVREL 2081
```

```
RESULT 7
MMWVRN
nonstructural polyprotein - rubella virus (strain Thertlen)
```

N:Contains: nonstructural protein NS1; nonstructural protein NS2; nonstructural prote
C:Species: rubella virus
C:Date: 30-Sep-1989 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
C:Accession: A35320; A29811
R:Domiguez, G.; Mang, C.Y.; Frey, T.K.
Virology 177, 225-238, 1990
A:Title: Sequence of the genome RNA of rubella virus: evidence for genetic rearrangem
A:Reference number: A35320; MUID:90281585
A:Molecule type: genomic RNA
A:Accession: A35320
A:Residues: 1-2205 <DOM>
A:Cross-references: GB:M15240; NID:9333971; PIDN:AA88528.1; PID:9333972
R:Frey, T.K.; Marti, L.D.
Gene 62, 85-99, 1988
A:Title: Sequence of the region coding for virion proteins C and E2 and the carboxy t
A:Reference number: A29811; MUID:88226020
A:Accession: A29811
A:Molecule type: genomic RNA
A:Residues: 1737-2205 <FRE>
A:Cross-references: GB:M15240
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: rubella virus nonstructural polyprotein
C:Keywords: nonstructural protein

Query Match 4.2%; Score 374.5; DB 1; Length 2205;
Best Local Similarity 19.2%; Pred. No. 6,3e-15;
Matches 454; Conservative 201; Mismatches 714; Indels 995; Gaps 104;

```
QY 3 ITTAIEQAAALAAANSALNAVAVVPPFLSKVQTEIILINMOPROLYPREVLNHPQRI 62
| : : : |
Db 40 VYTAQKRAIV-----AVIPRVETQMOVS-----DHP---AL 69
QY 63 HNELDYCARAGRCLEVGAPRS-----INDPNVLIHRCFLRPVQDYQWYS----- 111
| : : : |
Db 70 H-AISRTYR---HWLEMG--PREALHYLIDSPGL-----LREAVYKRWYALCLHRT 118
QY 112 -----APTRGPAAN-----GRRSALRLP-----ADRTYC-FDGF 141
| : : : |
Db 119 ARKLATALAEFSAEMAHAYVC---ALRGAPSGEYVHREDVPHGRAVADRLCYTTPM 175
QY 142 SRCFAAETGVALYSIHDIWPDVA-----E 167
| : : : |
Db 176 QMCELMRTIDATLLVAVDDMPVALAANVGDMDDLGIAHNLHDGGCPADCRGAGPTP 235
QY 168 AARHGXTFLYALAHLPREVLVLPCTYHTTSYLLIHDSGRAVVTYGDTSAGYNHDVSL 227
| : : : |
Db 236 GYTRPCTTYIYVL---PRTAHNGRLYRCGPRLMWTRDCAVAELSWVAOHCQO----- 286
QY 228 RAMIRTTKIVGHPVLEVRRAIGCHFVLLLTAAPEPSMPYVVPYRSTEVYVRSIFGPG 287
| : : : |
Db 287 -----ANVRAVRCITLPIRHVRSIQPS-----ARVR----- 311
QY 288 GSPSLPSPSAGSKSTFHAVPVHIMDRMLFGATLDDQAFCSRLMTYLGISKYTVGL 347
| : : : |
Db 312 -LPDVLVHLAEVRMWRVSLPRPVFOMLSYCTLSPDAYUSEVFEFKNALCHSTLACN 370
QY 348 VANEGWN---ASEDALTATTAAYLTICQRYLRQALISKMBRGHGVENAOKEFTRLYS- 403
| : : : |
Db 371 VLQEGKGTCAEDALCA-----VFAFRAMOSNARLAIMGAKCAADSLSV 417
QY 404 ---WLEFSGRDYIPGRLOFYAQCRRMLSA-----GFHLDPRVLEDESVPICRTPE 453
| : : : |
Db 418 AGML-----DPI-----WDAIKRFLGSVPLAEHMEEMEDDAVAARD----- 454
QY 454 LKKVAGKFCCEFMKWLQECTCFLEPAEGLVGDHGDNEAYEGSEVDPAEPAHLDVSGTYA 513
| : : : |
Db 455 -----RGLLEDGGRHLDITVQ-----PPKSPRPPEIATWI 484
QY 514 VGHQLE---ALYRALNVPOD-----IAARS 537
| : : : |
Db 485 VHAASEDRHCACAPRCVDYRERPPSAPAGOPDDEALIPWLFAERRALRCREWDFEALRAR 544
```

OY	538	RLIATVELVASPDRLCECRYLV-----GNKTFRTTVVDGAHLLEANGPEOY	581
Db	545	ADTAAAPARPARPRARYPTVLVYNRPAHNGPMTLDEPGEADAALVLCPLGOLPRGERH	604
OY	582	VLSFDAOSMGSGSHSLTYELTPALO--VWISSGLDCTATFFPGGAPBSAAGEVAAF	639
Db	605	-----FAGAHNCA--QARLOAFVHV-----VPPERPMADGG--ARA	639
OY	640	CSALRYNRFTORHSITLGLMLHPEDCL----IFFPPSPGHIMESANPCGEGITYTRT	695
Db	640	WAKFFRCSCMAAQR-----LDBEPVVMHLPTYDGVOPLI-----ALAIRT	679
OY	696	WSTISGESSDFS---PPEAAPAMAAT-----PGLP-----HS	724
Db	680	LAOGAALATALSVBDLPGCALFDANAVTAAVRAPROSAASPDPDPPPRRARSQRS	739
OY	725	-----TPVPSDIWVLPBPSEFOVDAPVPAPDPDAGLPQPVULTP PPPPHNK-----	773
Db	740	DARCTPPAPARDPPP-----APSPPAPRAGDPPV--PIPGAPDRARDAELE	787
OY	774	-----PSIPPPSRNR-----RLTTYPDGAKVYAGSLFESDCML	808
Db	788	VACPSGPPTSTRADDDSDIVESYARAACPVLHVRIDMDPPGCKV-----V	835
OY	809	VNASPNCHRPGGCLCHAFTQREPEAFYPTEFTIMEGLAA--YTLP RP-----	854
Db	836	VNAMEBELLASGVCGCALIFENATA--LANCRRLAPCPTGEAVATPGHC	884
OY	855	-----IIHAVAP--DYREONPKYLEAARE--TCSRGTNAATPYLGSGIV--QY	898
Db	885	GYNHILHAVARRRRROPALALEGBALLERAYRSIVALAARRMACVACPDLGAGYGMSA	944
OY	899	PVLSFPDAMERNHRPGDELTV-----TERPANFE	928
Db	945	AESIRALATATREPERYERSLICHCHDRATLTINASIYAGILAARVSPPTEPLAS--CP	1003
OY	929	ANKPAQVULTTTEDTARTANLALEDATVEGRACACTIS-----	969
Db	1004	ACDPGRPAORSASPPATPLG--DATAPEBCGCGCELCTRYTBTVDRAVYNLMLERD	1058
OY	970	-----PGIV-----HYOF-----RVVIDEAP	1037
Db	1059	RGATSWAMRIPEVVUYGREHLATHPRLNHYSVKRAEVRRPGMCGSDMMRCGMHGMPQ	1118
OY	978	-----TAGVPGSGKSRSIQOGDDVVUVVYPTRELNN--SMRRRGFAFTPTHAA	1023
Db	1119	VRCRPSNAHALCRTGVPRASTRG--GELDPTNCMYLRANAIVAQAARAGATYSACCP	1175
OY	1024	RVTIGR-----RVVIDEAP	1037
Db	1176	KCAAGRALSSEARTHEDFALSQWSASHADSPDGTGPLDTMETGVACACSRYWGSEH	1235
OY	1038	SLRPHLLLMORA-----SSVHLL--GPPNOI--PAL--	1066
Db	1236	EAPDHLLVLSHRAPNGPWGVLEVBARPEGNPIGHFCAGVGGBRRRSDBRPHMLA VP	1295
OY	1067	-----DFE-----HAGLVPAIRPELATPTS--WMXVT--	1090
Db	1296	LRSRGGCATDECGLAQATYYDLEFYRLGDMDAMARALASVGRPKKGYNIRWNMAAGA	1355
OY	1091	-----HRCRA--DYCELIR-----GAYPKI--	1108
Db	1356	GKTTRILAAFTREDLYVCPTNALNLHEIOAKLRABDIDIKNATYERRLLTKPLA V RRIY	1415
OY	1109	-----QTSRYLV-----SSLF-WNEP-	1123
Db	1416	DEAFTLGEYCASFVASTTAELVICVGRDQCGPHYANNCRTEVPDPWRPETERSRHTRFPD	1475
OY	1124	-----AIQOKLVHQAKKANPGATITHE	1147
Db	1476	CWAARLAGLDYDIERGEERTCFACJLMDGROYVDLHLASRETYV--RRLHEAGIRATYVRE	1533
OY	1148	AOGATFETTITATADARKL-IOSSRAHIAVALTRHTEKCVILD-APGLREVGISDIVY	1205

Db 1534 AAGMVGATACIHHVGRDGTDAVALATRLDIAIVSLTRASDALTLYHELEDGSLRAAGLS----- 1589

QY 1206 NNFFLAGEV-----GXHR-----PSVIRPNPNQNGTLOAFSPSCGISAYHOL 1250

Db 1590 --AFIDAGLAELEKVEPAIDIVVAEQAPPLP--PADGIPREADOVDPFCRTLEELV 1644

QY 1251 AEEELHRAAPVAALVPCPELEQGLLYMPEQELTVSDSVVFELTDLVHCRMAAPSORKAV 1310

Db 1645 FGRACH---PHRADLNRYTEGEREVRVM---RISHLLNKNTHEH-----PQNER-V 1689

QY 1311 LSTLVGRGRRTKLYEAHH--SDVRESLARFTP---TGVPQVAT-----CELYEL 1356

Db 1690 LSAVCA-----VRRYRAGEDSGTLRTAAVAHOPHRPRFOIIPPRVTAGVAOEWMMTYLRER 1744

QY 1357 VEAMWEKGGDSAVLEL-----DLCNRDVSRITFEOKXCNKF-----T 1394

Db 1745 IDLTVYVYOMGAANEELDTRVARRRPELPIAGCKTQASLSVPAFLKATLKCVDAALGPRDT 1804

QY 1395 TGETIANGKVGOGISAMSKTFCALFGPMWRATEKEILLPNIIFGYDAVEESVFMAAVS 1454

Db 1805 ECHHAAGCAGAGLEIRAMAEWQVMSPHRALQKTIIMRALRQGLVAAGHTEDEVDAWMO 1864

QY 1455 GAGSCMTFENDPSEFSDSTONFS-----IGLECVMEECGMPOMLIRLYLVMS 1503

Db 1865 AHYTTNAIEVDPTEDPTEDMNQTLTRQVLELISALGLPCA--ED-----YRALRA 1912

QY 1504 -AMIIQAPRESIKGFWKKHSGEGTLLMNTVNMMAIIACHYEPR--DFRYAA--FKGDD 1557

Db 1913 GSYCYLRELGSITEGCERTSGEPA TLNHTT-----VAACMMAMRVPKVRWAGITQGD 1967

QY 1558 SVVLCSDYRQSHNAAL-----IAGCGLKLVDRPIGLYAGVVAAPGLTLPDVR 1609

Db 1968 MVIIFLEPGARS--AALKMTPAEVGLFGFHHIRVKNHSTPTSPCG--HVGTAGLFLDVHM 2023

QY 1610 FAGRLSEKNMGFPERAEQQLAVCDFLRGLTNNVAQCVVVSRVYGVSPGLVHNLIGML 1669

Db 2024 QAIKVLCRFF--DPVULEEQVALLDRLRG-----VYAAPL----- 2057

QY 1670 OTIADGKAHFTETIKRPVLDLNTSI 1693

Db 2058 DTVAANAAAYDYSAEKVALIVREL 2081

RESULT 8

A38196

non-structural hypervariable probable nucleotide binding protein - hepatitis E virus

C:Species: hepatitis E virus

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-Feb-1997

C:Accession: A38196

C:Releasev. S.A.: Emerson, S.U.: Reyes, G.R.: Tsareva, T.S.: Legters, L.J.: Malik, I.A

Proc. Natl. Acad. Sci. U.S.A. 89, 559-563, 1992

A:Title: Characterization of a prototype strain of hepatitis E virus.

A:Reference number: A38196; MUID:92115700

A:Accession: A38196

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-141 <TSA>

A:Experimental source: strain SAR-55

A:Note: sequence extracted from NCBI Backbone (NCBIN:75998, NCBIp:76003)

C:Superfamily: hepatitis E virus nonstructural protein

C:Keywords: ATP

Query Match 4.1%; Score 372; DB 2; Length 141;

Best Local Similarity 50.3%; Pred. No. 1.7e-16;

Matches 84; Conservative 12; Mismatches 29; Indels 42; Gaps 7;

QY 651 ORHSLTGGUIMHPEGILGTFPPSPGHIMESANPFCGEGTLYTRIMS-----TS 699

Db 2 QRLSLTGNWFPEPGLGFAFPSPGHVHESANPFCGEGTLYTRIMSEVDAAVPSAPQD 61

700 GRSDFSPPEEAAPMAAATPGLPNSTRPVSDIWLVLPPSEEFQVDAAPV--PPADPA--GL 757

Db 62 GFS-----EFSIPSRAP-----TP-----AAPPAPADSPPTL 93
OY 758 PCGVVLTTPPPPPVHKKSIP-PPSRNRRLTYTPDGAKYAGSLFSS 803
Db 94 SAPAGEPAPGATAPARALITHOTAKRRRLFTYPDGSKYVAGSLFSS 140
RESULT 9
S01956
hypothetical protein, 195k - turnip yellow mosaic virus
C:Species: turnip yellow mosaic virus, TYMV
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 18-Jun-1999
C:Accession: S01956
R:Morch, M.D.; Boyer, J.C.; Haenni, A.L.
Nucleic Acids Res 16, 6157-6173, 1988
A:Title: Overlapping open reading frames revealed by complete nucleotide sequencing of
A:Reference number: S01955; M01D:88289359
A:Accession: S01956
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-1844 <MOR>
A:Cross-References: EMBL:X07441; NID:962222; PIDN:CAA30322.1; PID:962224
C:Superfamily: egplant mosaic virus RNA-directed RNA polymerase
Query Match 3.9%; Score 349; DB 2; Length 1844;
Best Local Similarity 20.6%; Pred. No. 1.9e-13;
Matches 385; Conservative 203; Mismatches 649; Indels 634; Gaps 93;
OY 157 LMD---LMPADVAEMARH-GXRLYAALHLPEVLLPFG-----YVHTSYLLIHD 204
Db 141 MMDALMYHPSQIMDLFLKRPNLERLYASLVPRPAHLSDQSFYKLYTTTRHTLHY- 199
OY 205 GBRVVTYGGDTSAGYNHDSILKARITRTKT-VGDHPLYTEVKAIG-CHFYLLLTAP 262
Db 200 -----VPGHENGASYNQ-SDAHSWLRLNSIRLGNHLSVTLLESMPVHSLIORGP 252
OY 263 EPSMPYVYPRSTEVYVSIFGP----- 286
Db 253 PDPBLSQAPPTLMTSLDFSSQYERLDVVSFRIPDAIELPQATFLQPLRDLVPPAVYN 312
OY 287 -----GGSPSLFPPSACSTKSTFHAVPVHIMRLMFGATLDDQAFCCSRLMT 333
Db 313 ALFTYRAVRLTSDPAFAFVRHMSKRPDHWTSNAMDVLQTF-----ALLNV 361
OY 334 YLR-GISYVTVG-----ALVANEGMNASEDLTAKITAYLTICQRYLRTQATISKMR 387
Db 362 PLRPNVVYHVLQSPILASLSLYLRQHMRRL-TATAVPLISFTLL-QRFV----- 408
OY 388 RLGEVHAQ-KETTRLYSWLFEKSGRDYIPGROLQFYAOCRRMLSAQFHLDPVLFVDES 446
Db 409 PLRPLAEVKSITAFRELRYRKE-----FHHPLDV-----FHLQHRKRWHSAT 453
OY 447 PCRCRTFLKRVAGKCCCFMRMLGECTCELEPAEGLVGDHNDNEAYSEVDPAPPAHL 506
Db 454 ----- 511
OY 507 DVSCTAAVHGQLEALYRALNVPQDIAARA-----SRLTAYELVASPR- 551
Db 465 KLP-----HALQKALLRLPISPLTATPFERSEQSKMLPNALSLWTLKRFALPQASL 519
OY 552 -----LECRVLGKTRTIVDGAHLAANGPDEQVLSFPAOSROSGASHLYELTPAG 607
Db 520 VLLALSESSILHKLFSPTLQAH-----DTYHNLHPGSSLSIOWERTP- 564
OY 608 LQVRISSNGL---DCTATFPGGAPSAAPGEVAAFCSALYRNRFTQHSILTGMLHPE 664
Db 565 LSIPTTALPFTPTSTAPDSEASLP---PAFAST-----FVPR----- 603
OY 665 GLLGTFPP---FSPG---HIMESANPFCGEGTLTYTWTSGF-----S 702
Db 604 -----PPPAASSPGAQPPTTAAAPTPIEPT-QRTHONSDLALLESSTSTPPPPIRS 655

OY 703 SDFSP-----PEAAAPAMAATPGL-PHSTPPVSDIWLPPPSSE----- 740
Db 656 PDMTPSAPVLFPIINSRRRPPQLPATPDLEPAHTPPPLISIPQDDPDSADMLGSHLLH 715
OY 741 FOYDAAPVPPAPDAGLPGCVVLTTPPPPPV----- 775
Db 716 HSLPAPPTPPLPSSQLLPALVLTNDPTAIGVLPFEELHPRRYENTATFLTRLRLSLEPSNH 775
OY 776 IPPSRNRRLTYTPDGAKYAGSLFSDODMLVNA-SNPGHRPG-----GLCH- 824
Db 776 LPQPTLN-CLLSAVSDQFKVSEELHESLQTLIPDSOLSAEETNVLGLSTENHTALHLX 834
OY 825 -----AFYQFPEAFYPTFERIMREGLAAYLTFRPIIHAVAPDYVEONKRLLEAAVRET 879
Db 835 NQOATVYSDRGPIILFGPSDITK-----IDITHTGPPSHSPG-KRLIGSQP- 881
OY 880 CSRRGTAAVPLLG-----SGLYQVVSLSFDAMERNHRGDELY-LTEPAAMFEA 929
Db 882 -SAKGHPDPLIRAMKSFKVSQNY-LPFS-----EAHNHPTISIAKNLISMKNKGFQD 933
OY 930 NKRAQVLTTEBTARTANALEIDATEVGRACAGCTISPC-----IYHQFAGVPGSG 985
Db 934 VLSLDVSTGQRTGTPPKERILQIDHYLDT-----NPGKTPVYHF--AGFAGCG 981
OY 986 KRSRIGQ-----GDVVVYVPTRELNSMR-RCFAA-FTPHTAARTIGRRV 1032
Db 982 KTYPIQOLKTKLFRKFR-VSCETTELRTKMTAMELHSGSQRFTWESSILKSSRLIV 1040
OY 1033 IDEAPSLPRLHLLHMQRASSVHL--LQDP-----NOIPAIDEHAGLV 1075
Db 1041 IDEIYKMPKGYDLSTILAPALFELVILGDPLGEVHSGSKDSNNHRLPS-ETRLRLP 1097
OY 1076 ATRPELAPTSWMXVTHRCRADVCEL-RCAYPRIG-TRSVLRSLF 1119
Db 1098 YI---DMTCW--STRIPQCLARLPQIHSEFNAQVIGSVTPPHDOSPVLTSHASSLT 1151
OY 1120 WNEPAIGOKLVYQAAKANPAGATFVEAGATFETITATDANGLIQSSRAHAIVAL 1179
Db 1152 FN-SIGYR-----SCTISSOGGLFCPCPAIIVLDNTRYKWLSS-ANGVAL 1194
OY 1180 TR-----HTEKCVILDA--PCLREVGISDYVNNFLL 1210
Db 1195 TRSRSGVQFPGPSSVYGVGNGSSAMFSDAFNNSLIIMDRFPSPFLQOL-KLITSPFLT 1251
OY 1211 AGEVEGXHRPS-----VIRGNPDQNLGLOAFPPS 1241
Db 1252 RGPRLNGAPPSASPTHRSPFHLRPHIPLSYDRDEYTVNFTLPDQGETRLDT-HFLRP 1309
OY 1242 COISAVHQLAEELGHRPAPVAAVLPCCPELEQGLLYMPOELIYVSDVLEL-----T 1294
Db 1310 SRLPLHFDLPRAL-TPPPYSTVDP-POAKAPVY-PGFF-FDSLAAFFLPAHDPSTR 1363
OY 1295 DIVHCMAAPS--ORKAVLSTLVGRYGRRTKLYEAAH-----SDVRESLARET 1340
Db 1364 EILHKDOSSNQPFWRPESLSC-----QPSLLISAKHAPNHDPPLLPAISIKRR-RFR 1416
OY 1341 PTIGPYOAT------C-----ELYELVAMEYKGGDSAYLE 1372
Db 1417 PSDSPHOITADADVVLGLQLFHSICRAYSRQPNSTVPPNBELECIISLNEYAOLSKTOS 1476
OY 1373 LDCNRDVS-----RITFEKXCNKFTTGETIAGKVGQGISAMSKFCALEGF--W 1422
Db 1477 TIVANASRSDPWRHTTVKIFKAQOKKVDNGSIFGSMKACQITLALHNDVILVIGVYKK 1536
OY 1423 FRAIEKEILLAPNIF-----YGAYESVFAAAYSGAGSCMVENDFSEPDSTQ 1473
Db 1537 ORIFDN-ADRPNIYSHCGKTPNOLRDMCOEHL-----THSTPKIANDYTADPOG 1586
OY 1474 NNFSLGLECVYMEEGMPQMLIRLY-HIVRSANITLAPKESLKGFMKKHSGEGTLMMNT 1532
Db 1587 HGESVLEALMKMKRLNIPSHLIQHLVHLKTNVSTQPGPLTCM-----RLTGEPTYDDNT 1641


```

QY 1382 ---HTEFOKXCKNFETGETIANGXGGOGISAMSKFCLPDP---WPALEKEITALLP 1435
Db 1490 RHTTVKIFAKQKHVNDGSIETFGSWKACQTLALMHDIYLLVLPQVKKIQRIDFN---ADRP 1546
QY 1436 PNIF-----YDAYEESEVFAAASGAGSCWFEENDESEFSDTONNKSIGLECYVME 1486
Db 1547 PNTHSGCKTPNOLRDMCQEHLE-----THSTPKIANDYTAFDQSQHESSVLEALCKMK 1599
QY 1487 ECGHPQULIRLY-HLVASAIITLQAPKESLKGFKKHSGEPGLLNTYNNWMAIILHCEFF 1545
Db 1600 RLNIPLSHLIQILHNLKLTNVSTGEPRLTCM-----RLTGEPRGYDDNTDNLAVIYSQYD 1654
QY 1546 RDEPVAFAFKGDDSVVLCSDY--ROSRNAALAIAGCILKAKVY-----1586
Db 1655 GSCPIWV-SGDDSLI---DHRLPTRHDPWSVLKRLRLRKLELTSHPLCCGYVGPACCI 1710
QY 1567 -RPIGLYAGVYVAPGCTLPD-----VWRF--AGRISEKNWGPGEBAQRLVACDF 1636
Db 1711 RNPILATCKILAVDDALDRLISVLTETPTGHLIGESIMHILDETHVQYOSACDF 1768

```



```

OY 650 TORHSLTGLMLPEGLLGPSPSPGHIMESANPFCGEGTIVTRTWSTSGFSDPSPE 709
Db 622 -----LQPTTTSKEBSLKT-----VSDDVESSIKE 648
OY 710 AARPAARATPGLPSTPPVSDIWLPPSEPOVDAVPAPDAPGLPQPVLLTPPPP 769
Db 649 VVRKSEISMGLGNT--VSEDEFORSTEIESLQ-----679
OY 770 PVKPPSLPPSPSRNRRLIYTPDGAKYVAGSLEFSDC---DWLVMSNPGHRPGGICHA 825
Db 680 QFMHVSSTETIIRKOM-----HAMVYTGPLKYQCKNYKNDISLV-----716
OY 826 FYGRPEAFPTPEFIRREGIAATLIPRPIIHAVAPDYREQPKLEAAYRETCSRGT 885
Db 717 -----ASLSAAVSNLKLITKDTA--AIDLETKERGVY-DVCLKMW 755
OY 886 A-----AYPLGSGIYQVPVSLSPDAMERNRHPGDELITPEPAWFEANKPAQPVLT 938
Db 756 VKPLSGHANGVMDSDYKCFVAL-----LYDGEN-----I 787
OY 939 ITEDTARTANLALEIDAATEVGRACAGCTI---SPGIYHYQPTA--GVPGSGKSRSIQ 992
Db 788 VCGETWRRAVAVSSESLVYSDMGKIRAIRSVLKDGEPHISSAKVTLVDGVPGCGTKETLS 847
OY 993 G---DYDVVVVPTRELNSRRR---GPAFTPHTAARY-----TIGR-----RVID 1034
Db 848 RVNFDDELVLVPGQAEMIRRRANSGLVAATKEVNTVDSFLMNGRPGCCYKRRFLD 907
OY 1035 EAPSLPHLL--LHMQRASSVHLGDPOIPAID---FEHAGLVPAIRBELAPTSWMX 1088
Db 908 EGLMLHPCGVNPLVGSLSCEAFVYGDTOQIPIYNVATPYPKHLSOLEVDAVETR--R 965
OY 1089 VTHRCPADVCELIRGAYP-KIQTTSHVLSLFW-----NEPAIGOKLVYTOAA 1135
Db 966 TLRCPADITFFLNQKYEGVCTSVTSRVSHEVILQAAVMNPYSKPLKGVITFTQSD 1025
OY 1136 KA-----ANPGATVHEAGATFTETIIT-ATDARGLIOSRAHAIVALTRHEK---- 1185
Db 1026 KSLLSRGYEDVHTVHEVOGETFEDVSLVNLPTPVGIISKOSPHLVSLSRHTRSIKYY 1085
OY 1186 CVLLDA--PGLLREVG--ISDVIYN-----NFLAGGEVGHHRPSYI 1223
Db 1086 TVVLDVAVSVLRDECVSSYLDMKVDSVTOOLOIESYKGVNLVFAAPKTG----- 1138
OY 1224 PRGNPDONLGTLOAFPPSQ-----ISAYHOLAEELGHRPAPVAVLPPCELEOGLLY 1277
Db 1139 -----DVSDMQYDYDKCLPGNSTILNEYDAYTMQI-----1168
OY 1278 MQEELTVSDSVLF-----ELTDIVHCHMAAPSOK--AVLSTLVGRYGR--1321
Db 1169 RENSILNVKDCVLDMSKSVPLPRESETTLKPVIRTAAEKPRKPGLENLVAMIKRNFNSPE 1228
OY 1322 -----TKLYEAHSDVRESLARFIPTIGPVQATTCELYELVEMVKEGOD 1366
Db 1229 LVGVVDIEDTASLVYDKFED--AYLKEKKPKKPIPLISRAS-----LEKWIKE--OE 1277
OY 1367 GSAVLEL-----DLGNRDVSRITPEOKXONKFTTGETIANGKVGOGISAMSKTECALFG 1420
Db 1278 KSTIGQLADFDLIDPAVQYRHMIKQPKORLDSIQTEYRAL-QTIYVHKKKINALFG 1336
OY 1421 PWFRAIEKEIILALPN--IFYGDAVEESY--FAAIVSAGSGMWFEENDSEFDSTQNNF 1476
Db 1337 PVFSELTROLLETIDSSRPMFYTRKPTQIEEFPSDLSNVPMIDLIELDISKYDKSQNEF 1396
OY 1477 SLGLECVVNEEGCMPQWLRLYHVSAMILQAPRESLKG-----WKHSGEPGT 1527
Db 1397 HCAVEYEIKRIGLDLEAEV-----WKHGRKTTLLKYTAGIKTCLMWQKSGDVTT 1449
OY 1528 LLMNTVWMMNAIIAHCEYFRDFRVAAFKGGDSVYLC---SDYRQSRNAALIAGCGLKLKY 1584
Db 1450 FIGNITIIAAGTSSMLPMERLLKGAFCGDDSLIFPKGTDFPDIOGANLL--WNEFAKL 1507

```

```

OY 1585 DYRPIGLXGVVVA-----PGJSTLPDYYVRFAGRLSEKNNWGGPERARQOLRLAVCDFLRGLT 1641
Db 1508 FRKRYGFCGRYIIHHDRGCIYVYDPLKLSLGAHNI--KNRHELEFRTSLCDVAGSLN 1566
OY 1642 NVAOV--CVDVVSRYVGVSP 1659
Db 1567 NCAYTHINDAVEVITKTAP 1586

```

Search completed: March 7, 2002, 14:04:18
Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:03:52 ; Search time 16.68 Seconds
(without alignments)
3732.427 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013

Sequence: 1 PGTTTAEQAALAAANSALA.....FTETIKPYLDITNSIQHVE 1698

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7459	82.8	1693	1 POLN_HEVPA	P33424 hepatitis e
2	7448	82.6	1693	1 POLN_HEVBU	P29324 hepatitis e
3	7433	82.5	1691	1 POLN_HEVME	003495 hepatitis e
4	7433	82.5	1693	1 POLN_HEVNY	004610 hepatitis e
5	406.5	4.5	1648	1 RRPD_GCMVS	P19523 cucumber gr
6	374.5	4.2	2205	1 POLN_RUBVT	P13889 rubella vir
7	351.5	3.9	1612	1 RRPD_ORSVS	084133 odontogloss
8	349	3.8	1844	1 POLR_TYMC	P10358 turnip yell
9	346	3.8	1844	1 RRPD_TYMC	P28477 turnip yell
10	339.5	3.8	1612	1 RRPD_ORSYC	P86659 odontogloss
11	322.5	3.6	1616	1 RRPD_TOML	P03387 tomato mosa
12	321.5	3.6	1616	1 RRPD_TOMK1	094146 tomato mosa
13	320.5	3.6	1616	1 RRPD_TOMK1	094146 tomato mosa
14	320	3.6	1597	1 RRPD_CRMV	066220 chinese rap
15	319.5	3.5	1616	1 RRPD_TMOB	P90211 tobamovirus
16	316.5	3.5	1616	1 RRPD_TOMK2	P86676 tomato mosa
17	313	3.5	1601	1 RRPD_TVCV	088320 turnip yell
18	312.5	3.5	1612	1 RRPD_PMAVJ	P89657 pepper mild
19	312.5	3.5	1612	1 RRPD_PMAVJ	P29098 pepper mild
20	311	3.5	1748	1 POLR_ELV	P35228 erwinium la
21	310	3.4	1844	1 POLR_TYMA	P20128 turnip yell
22	305.5	3.4	1776	1 POLR_OTMV	P20127 onion yell
23	301.5	3.3	1616	1 RRPD_TWVM	093058 tobacco mos
24	299.5	3.3	1616	1 RRPD_TWVB	093058 tobacco mos
25	295.5	3.3	1616	1 RRPD_TWV	P03386 tobacco mos
26	293.5	3.3	1616	1 RRPD_TWVKR	P30738 tobacco mos
27	293	3.3	1874	1 POLR_KYMAJ	P36304 Kennedy ye
28	291.5	3.2	1616	1 RRPD_TWVA	098745 tobacco mos
29	287.5	3.2	1839	1 POLR_EPMV	P20126 eggplant mo
30	261.5	2.9	1609	1 RRPD_TWGMV	P18333 tobacco mll
31	254	2.8	1629	1 RRPD_SHMV	P89202 sunn-hemp m
32	243.5	2.7	1707	1 194K_TPVSY	P05080 tobacco rat
33	243.5	2.7	1884	1 RRPD_ACTLSP	P27738 apple chlor

34	242.5	2.7	1885	1 RRPD_ACTLSA	P54891 apple chlor
35	218	2.4	2512	1 POLN_STNDV	P03317 sindbis vir
36	211	2.3	2514	1 POLN_STNDV	P27283 sindbis vir
37	207	2.3	2514	1 POLN_ONNVG	P13886 O'nyong-nyo
38	201.5	2.2	1718	1 RRPD_SHVX	004575 shallot vir
39	200	2.2	2431	1 POLN_SFV	P08411 semliki for
40	194	2.2	1967	1 RRPD_PVMR	P17965 potato viru
41	194	2.2	2485	1 POLN_EEVV3	P36327 venezuelan
42	184.5	2.0	3175	1 RPOA_EAV	P19811 equine arte
43	177.5	2.0	2492	1 POLN_EEVT	P27282 venezuelan
44	176	2.0	1385	1 RRPD_PIAMV	007518 plantago as
45	174.5	1.9	2492	1 POLN_EEVP	P36328 venezuelan

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	1693 AA.
ID	POLN_HEVPA			
AC	P33424;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE			
DE	(EC 2.7.7.48); HELICASE].			
OS	Hepatitis E virus (strain Pakistan) (HEV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage.			
OX	NCBI_TaxID=33774;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92115700; PubMed=1731327;			
RA	Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,			
RA	Malik I.A., Iqbal M., Purcell R.H.;			
RT	"Characterization of a prototype strain of hepatitis E virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).			
CC	-I- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF			
CC	ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M80581; AAA45725.1; -			
DR	MEROPS: C41.001; -			
DR	InterPro: IPR002589; DUF27.			
DR	InterPro: IPR002588; V_methyltransf.			
DR	InterPro: IPR000606; Viral_helisel.			
DR	Pfam: PF01661; DUF27; 1.			
DR	Pfam: PF01443; Viral_helisel; 1.			
DR	Pfam: PF01660; Vmethyltransf; 1.			
DR	SMART: SM00506; Atpp; 1.			
KW	Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;			
KW	ATP-binding.			
FT	NE-BIND 975 982 ATP (POTENTIAL).			
SO	SEQUENCE 1693 AA; 185149 MW; 5A0F03FB1F99E8 CRC64;			
Query Match	82.8%; Score 7459; DB 1; Length 1693;			
Best Local Similarity	81.8%; Pred. No. 0;			
Matches 1401; Conservative 109; Mismatches 160; Indels 42; Gaps 7;				
QY	1 PGTTTAEQAALAAANSALANAVVVPFLSRVOTETIILINMOPROLVFPEVIMNPIOR 60			
DB	10 PGTTTAEQAALAAANSALANAVVVPFLSHQOIEIILINMOPROLVFPEVIMNPIOR 69			
QY	61 VHNLELYCARAGRCLEVGAPRSTINDPNVLRHCFELAPVDVORVWYSAETRGPAAN 120			
DB	70 VHNLELYCARAGRCLEVGAPRSTINDPNVLRHCFELAPVDVORVWYSAETRGPAAN 129			

```

121 CRRSALRGLPADRTVCDFGSCAFACAEETGVALYSLHDLMPADVAEAMARHCXTRLYAA 180
130 CRRSALRGLPADRTVCDFGSCAFACAEETGVALYSLHDLMPADVAEAMARHCXTRLYAA 189
181 LHLPEVLLPGTYHTTYSLLIHDGDAVYVEGDTSAQYNDVSLIRAMIRTTKIVGDH 240
190 LHLPEVLLPGTYHTTYSLLIHDGDAVYVEGDTSAQYNDVSLIRAMIRTTKIVGDH 249
241 PLVIEVRAICGHFVLLLAAPERSPMYVYPRSTREVVYRSLFEGGSPSLFSPACSTRK 300
250 PLVIEVRAICGHFVLLLAAPERSPMYVYPRSTREVVYRSLFEGGSPSLFSPACSTRK 309
301 STEHAPVIMIRMLFAGATLDDOAFCCSRMLTYLIGISYKTYGVALVNEGNASEDL 360
310 STEHAPVIMIRMLFAGATLDDOAFCCSRMLTYLIGISYKTYGVALVNEGNASEDL 369
361 TAXITAAVLTJCHORYLRTQALSKGRRLGVEHAQKFTIRLYSMLFEKSGRDYIPGRQ 420
370 TAXITAAVLTJCHORYLRTQALSKGRRLGVEHAQKFTIRLYSMLFEKSGRDYIPGRQ 429
421 FTAQCRMLISAGFHLDRVLVFEDESPCRCRTFLKVAACKFCCFMWLODECCFLERAP 480
430 FTAQCRMLISAGFHLDRVLVFEDESAPCHCIRTAIRAVSKFCCFMWLODECCFLERAP 489
481 GVGGDHNDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNVPDODIARASRL 540
490 GVGGDHNDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNVPDODIARASRL 549
541 ATVEVLASDRLECRVYLNKTFRTTYVDGAHLLEANGPREQYVLSFSPASQSKAGSHSLT 600
550 ATVEVLASDRLECRVYLNKTFRTTYVDGAHLLEANGPREQYVLSFSPASQSKAGSHSLT 609
601 YELTPAGLOVRISNGLDCTATPPPGGAPSAAPGEVAACSAALYRNRTOHRSLSLGLM 660
610 YASASAGLEVRVYAGLDRAVAFAPGVSPRASAPGEVTAFCASALYRNRTOHRSLSLGLM 669
661 LHPGELGLTFPPSPGHIMESANPFCGEGLYTRTWS-----TSFGSSDFSPPE 709
670 LHPGELGLTFPPSPGHIMESANPFCGEGLYTRTWS-----TSFGSSDFSPPE 724
710 AAAPMAATPPGLPHSTPPYSDIWLPPRSEEROVDAAPV-PRAPDA-GLPGFVLTTPPP 767
725 PSLPSRAAP-----PP-----AAAPMAATPPGLPHSTPPYSDIWLPPRSEEROVDAAPV-PRAPDA-GLPGFVLTTPPP 761
768 PPPVHAPSTP-PPSRNRRLTYTPDCAKYVAGSLFESDDMLVNAANPCHRGCGCHAF 826
768 GATAPAPALTTHOTARHRLTLTYPDGSKVAFAGSLFESTCTWLNVASVNDHRRGGLCHAF 821
827 YORFPEAFYPTFEIMREGIAAYTLPPRLIHAVADRYEONPKRLEAAVRETCSRRTA 886
822 YORFPEAFYPTFEIMREGIAAYTLPPRLIHAVADRYEONPKRLEAAVRETCSRRTA 881
887 APPLUGSGIYQVVSLSFPAEMERNHRRGDELITYTERPAANWFEANPAQOVLITTEBTANT 946
882 APPLUGSGIYQVVSLSFPAEMERNHRRGDELITYTERPAANWFEANPAQOVLITTEBTANT 941
947 ANLALEIDATEVGRACACCTISPGIVHVOFTAGVGSKSRSIOGDVDVYVPPRELR 1006
942 ANLALEIDATEVGRACACCTISPGIVHVOFTAGVGSKSRSIOGDVDVYVPPRELR 1001
1007 NSMRKRGFAATPHTAARTIGRRVVIDEAPSLPHLLILHMQRASSVLLGDPNQIPAI 1066
1002 NSMRKRGFAATPHTAARTIGRRVVIDEAPSLPHLLILHMQRASSVLLGDPNQIPAI 1061
1067 DEHAGLVAPRPELAPTSWMXVTHRCPADVCELLINGAVPKIQTSRVLSLFWNPATG 1126
1062 DEHAGLVAPRPELAPTSWMXVTHRCPADVCELLINGAVPKIQTSRVLSLFWNPATG 1121
1127 OKLVYQAANKANPGAITVHEOGAFTFTTTIATADAGLIQSSNAHAIVALTTRTEKC 1186
1122 OKLVYQAANKANPGAITVHEOGAFTFTTTIATADAGLIQSSNAHAIVALTTRTEKC 1181

```

```

1187 VILDAAGLLEVEGIDSPVIVNNEFLAGGEVXHRPSYIPRGNPDONLGLQAPPPSCQISA 1246
1182 VILDAAGLLEVEGIDSPVIVNNEFLAGGEVXHRPSYIPRGNPDONLGLQAPPPSCQISA 1241
1247 YHQLAEELGHRAPAAVULPPCELEQGLLYMPQELTYSDSVLVEELDVIYCRMAAPSO 1306
1242 YHQLAEELGHRAPAAVULPPCELEQGLLYMPQELTYSDSVLVEELDVIYCRMAAPSO 1301
1307 KRAVISTLVGRYGRTKILEAASDVRESLAPFTIPIGVQATTCLELYELVAMEVEKGD 1366
1302 KRAVISTLVGRYGRTKILEAASDVRESLAPFTIPIGVQATTCLELYELVAMEVEKGD 1361
1367 GSAVLELDICNDVSKRTPEOKXCNKFTTGETIAGKVGQGISANSKTFICALFGWFRAT 1426
1362 GSAVLELDICNDVSKRTPEOKXCNKFTTGETIAGKVGQGISANSKTFICALFGWFRAT 1421
1427 EKEITALLPPIFYGDAYEESFPAASGACMPFENDFSEFDSTONNESLGLCYWHE 1486
1422 EKEITALLPPIFYGDAYEESFPAASGACMPFENDFSEFDSTONNESLGLCYWHE 1481
1487 ECGMPOMLIRLHLVRSAMIILOAPRESLKGFWKKHSGEPGTLIMNTVMNALIACYEF 1546
1482 ECGMPOMLIRLHLVRSAMIILOAPRESLKGFWKKHSGEPGTLIMNTVMNALIACYEF 1541
1547 DERVAAFKGDSDVSLCSDYRQSRNMAALTAGGLKLYDYRIGLYAGVVAAPGLGTLPD 1606
1542 DERVAAFKGDSDVSLCSDYRQSRNMAALTAGGLKLYDYRIGLYAGVVAAPGLGTLPD 1601
1607 VVRPAGRLSEKMWGCPREAPBOLRLACDEGLTNVAVOYDVYSRVYVSGPLVHNI 1666
1602 VVRPAGRLSEKMWGCPREAPBOLRLACDEGLTNVAVOYDVYSRVYVSGPLVHNI 1661
1667 GMLQITADGKAHFTETIKPVLDTLNSIIORE 1698
1662 GMLQITADGKAHFTETIKPVLDTLNSIIORE 1693

RESULT 2
POLN_HEVBU
ID POLN_HEVBU STANDARD: PRF: 1693 AA.
AC P29324:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE
(LOC 2.7.7.48); HELICASE].
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=31767;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RA Fry K.E., Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RL Virology 185:120-131(1991).
CC -I- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (EN-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M73218; AAA5734.1;
DR PIR: A40778; MNMWH.
DR MEROPS: C41.001;
DR InterPro: IPR002589; DUF27.
DR InterPro: IPR002588; V_methyltransf.

```

DR InterPro: IPRO00606; Viral_helcse1.
DR Pfam: PF01661; DUF27; 1.
DR Pfam: PF01443; Viral_helcse1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
DR SMART: SM00306; Altp; 1.
DR Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
KW ATP-binding.
FT NP_BIND 975 ATP (POTENTIAL).
EQ SEQUENCE 1693 AA; 185191 MW; 2P355E46BED219B CRC64;

Query Match	82.6%;	Score 7448;	DB 1;	length 1693;
Best Local Similarity	81.8%;	Pred. No. 0;		
Matches 1396;	Conservative 110;	Mismatches 169;	Indels 32;	Gaps 6

QY	1	PGITTAIEOAA	LAANS	LA	AAV	VP	RELS	VO	CE	LI	IN	MO	PO	LY	FE	RE	V	LN	HH	IO	R	60
Db	10	PGITTAIEOAA	LAANS	LA	AAV	VP	RELS	VO	CE	LI	IN	MO	PO	LY	FE	RE	V	LN	HH	IO	R	69
QY	61	VINHELD	OC	RA	AR	AG	RC	LE	VA	GH	AP	RS	IND	P	BN	V	A	H	R	C	E	120
Db	70	VINHELD	EL	YC	AS	RG	CE	LE	IG	AH	RS	IND	P	BN	V	A	H	R	C	E	129	
QY	121	CRSAL	RG	L	P	AD	RT	YC	P	D	G	F	S	C	A	F	A	E	T	G	V	180
Db	130	CRSAL	RG	L	P	AD	RT	YC	P	D	G	F	S	C	A	F	A	E	T	G	V	189
QY	181	LH	PE	V	L	L	P	G	T	H	T	S	Y	L	L	I	H	G	D	A	V	240
Db	190	LH	PE	V	L	L	P	G	T	H	T	S	Y	L	L	I	H	G	D	A	V	249
QY	241	PL	Y	E	R	A	I	C	H	F	V	L	L	T	A	P	E	S	P	M	P	300
Db	250	PL	Y	E	R	A	I	C	H	F	V	L	L	T	A	P	E	S	P	M	P	309
QY	301	ST	F	H	A	V	P	H	I	M	D	R	L	M	F	E	G	A	T	L	D	360
Db	310	ST	F	H	A	V	P	H	I	M	D	R	L	M	F	E	G	A	T	L	D	369
QY	361	T	A	X	I	T	A	V	L	T	C	H	O	R	Y	L	T	O	A	I	S	420
Db	370	T	A	V	I	T	A	V	L	T	C	H	O	R	Y	L	T	O	A	I	S	429
QY	421	F	Y	A	O	C	R	M	L	S	G	F	L	D	R	V	L	F	E	D	S	480
Db	430	F	Y	A	O	C	R	M	L	S	G	F	L	D	R	V	L	F	E	D	S	489
QY	481	G	L	V	D	H	G	H	D	N	E	A	V	G	S	E	V	D	P	A	E	540
Db	490	G	A	V	D	H	G	H	D	N	E	A	V	G	S	E	V	D	P	A	E	549
QY	541	A	T	V	E	L	V	A	S	P	D	L	E	R	T	V	L	G	N	T	F	600
Db	550	A	T	V	E	L	V	A	S	P	D	L	E	R	T	V	L	G	N	T	F	609
QY	601	Y	E	L	P	A	G	I	O	V	I	S	S	N	G	L	D	C	T	F	P	660
Db	610	Y	A	A	S	A	G	I	E	V	Y	A	A	G	L	D	H	R	A	V	F	669
QY	661	L	H	P	E	G	L	G	I	F	P	F	S	P	G	H	I	W	E	S	A	714
Db	670	F	H	P	E	G	L	G	I	F	P	F	S	P	G	H	I	W	E	S	A	719
QY	715	M	A	A	P	G	L	P	H	S	P	R	P	S	D	I	W	L	P	P	E	772
Db	730	R	A	A	P	P	L	-----	A	A	L	P	P	P	P	S	P	P	S	P	A	766
QY	773	K	P	S	T	P	-----	P	S	R	R	L	L	Y	T	P	D	G	A	K	Y	831
Db	767	A	P	A	T	H	O	T	A	R	R	R	L	L	F	T	Y	P	O	S	K	826
QY	832	E	A	F	P	T	E	I	M	R	E	G	I	A	A	T	L	P	R	I	A	891

Dd	827	ASPDAAEYVMDGAAATTLTPRPILIAHAVADRYLLENRPKLEAAARECISRLGTAAVLL	886
Qy	892	GGGIVYOVVSLSEDAWERNHRPGDELYLAEPAAMFEANKPAQPVLTITEDTARTANAL	951
Dd	887	GTGIVYVIGPSFDAMERNHRPGDELYLPETLAARFEANRPTRPVLITIEDVARTANLAI	946
Qy	952	EIDAAETVGRACAGCTISPEIVHGYOTFAGYPGSGKRSITQOGDVVYVYPTRELRNSRR	1011
Dd	947	ELDASATDVGACACCRATPGVVOYQETAGVPGSGKRSITQADVVDVYVYPTRELRNSRR	1006
Qy	1012	RGFAAFPHPTAARTYITRRRVIDEAASLPRLHLLIMQASSVHLLGDNPQIPALDFEHA	1071
Dd	1007	RGFAAFPHPTAARTYITRRRVIDEAASLPRLHLLIMQAAVYHLLGDNPQIPALDFEHA	1066
Qy	1072	GLVPAIRBELAPTSWMVAVYTRHRCBADVCCELLIGAYPKTIQTTSRVLSLFENEPALQKLVX	1131
Dd	1067	GLVPAIRBDLQPTSMVHVTHRWPADVCCELLIGAYPMIQTTSRVLSLTFWGEPAVQKLVF	1126
Qy	1132	TQAAKANPGATVYHEAGATFTETTTIATADARGLIOSSRAHAIVALTTRHTEKCVIIDA	1191
Dd	1127	TQAAKANPGATVYHEAGATFTETTTIATADARGLIOSSRAHAIVALTTRHTEKCVIIDA	1186
Qy	1192	PGLLRKRGISQVITYNNFELAGGEGVGHRRPVIIRGPNDDNIGTIAFPSPCOISAYHOLA	1251
Dd	1187	PGLLRKRGISQVITYNNFELAGGEGVGHRRPVIIRGPNDDNIGTIAFPSPCOISAFHOLA	1246
Qy	1252	ELGLHRPAPVAVLPPEPELEOGLIYMPQELTVSDVLVFEFLTDIYHCMAAPSORKAVL	1311
Dd	1247	ELGLHRPAPVAVLPPEPELEOGLIYMPQELTVSDVLVFEFLTDIYHCMAAPSORKAVL	1306
Qy	1312	STLVGRYGRRTKLYEAAHSDVRESLARFIPTIGPVQATTCELYELVEAMVEKGDSAVL	1371
Dd	1307	STLVGRYGRRTKLYNASHSDVRSRLARFIPTIGPVQATTCELYELVEAMVEKGDSAVL	1366
Qy	1372	EFLDCNRDYSKITEFQKXCKKFTTGELIANGKVGQGISASWKTFCALBEPMPRAIEKELL	1433
Dd	1367	EFLDCNRDYSKITEFQKXCKKFTTGELIANGKVGQGISASWKTFCALBEPMPRAIEKALL	1426
Qy	1432	ALLPNTPIFYGATYESYFAAASVAGSCAMFENDFSEFPSTONNFSLGEGCVMEECGMP	1491
Dd	1427	ALLPQGVYFGAIFDITYFSAVAAKARASWFENDFSEFPSTONNSLGLIECAIMECCMP	1486
Qy	1492	QMLIRLYHLVLSANILQAPKRESLKGFWKXHSBPGTLLMNTVWNNALIAHCYEPDFRYA	1551
Dd	1487	QMLIRLYHLVLSANILQAPKRESLKGFWKXHSBPGTLLMNTVWNNALVITHCYDFRDFQYA	1546
Qy	1552	AFKGDSDVYLCSDYRQSRNAALIAAGCGKLAKDYRPIGLIYAGVYVAPGLTLPVYRFA	1611
Dd	1547	AFKDDSDVYLCSDYRQSRNAALIAAGCGKLAKDYRPIGLIYAGVYVAPGLTLPVYRFA	1606
Qy	1612	GRLEKMNKGPPEPAEOLRLAVCDPLRGLNNVQVVDVAVSRVYGSPLVNLNLGMOT	1671
Dd	1607	GRLEKMNKGPPEPAEOLRLAVSDFLRKLTLNNAQMCVDAVSRVYGSPLVNLNLGMLOA	1666
Qy	1672	IADGKAHETETIKPELDLTNSIIQRYE	1698
Dd	1667	VADGKAHETESVKRPVLDLTNSIICRYE	1693
RESULT 3			
POLN_HEVME STANDARD; PRT; 1691 AA.			
ID	AC	003495;	
DT	DT	01-OCT-1993 (Rel. 27, Created)	
DT	DT	01-OCT-1993 (Rel. 27, Last sequence update)	
DT	DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	DE	NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE	
DE	DE	(EC 2.7.7.48); HELICASE].	
OS	OS	Hepatitis E virus (strain Mexico) (HEV).	
OC	OC	Virusess; ssRNA positive-strand viruses, no DNA stage.	
OK	OK	NCBI_taxid=31766;	
RN	RN	[1]	
RP	RP	SEQUENCE FROM N.A.	

RX MEDLINE-93079857; PubMed-1448913;
 RA Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,
 RA Bradley D.W., Tam A.W., Reyes G.R.,
 RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis
 RT E virus (HEV)."
 RT Virology 191:550-558(1992).
 RN
 RP SEQUENCE OF 965-1691 FROM N.A.
 RX MEDLINE-92271462; PubMed-1589964;
 RA Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,
 RA Platak M., Feldman R.A., Yun K.Y., Purdy M.A., McCusland K.A.,
 RA Bradley D.W., Reyes G.R.,
 RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
 RT region encoding consensus motifs for an RNA-dependent RNA polymerase
 RT and an ATP/GTP binding site."
 RT Virus Genes 6:173-185(1992).
 CC -I- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
 CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M74506; AAA45730.1;
 CC PIR: A44212; A44212.
 CC MEROPS: C41.001; -
 DR InterPro: IPR002589; DUF27.
 DR InterPro: IPR002588; V_methyltransf.
 DR InterPro: IPR000606; Viral_helices1.
 DR Pfam: PF01661; DUF27.1.
 DR Pfam: PF01443; Viral_helices1.1.
 DR Pfam: PF01660; Methyltransf.1.
 DR SMART: SM00506; Altp.1.
 DR Polyprotein: Transferase; RNA-directed RNA polymerase; Helicase;
 KW ATP-binding.
 FT NP_BIND 973 980 ATP (POTENTIAL).
 FT SEQUENCE 1691 AA; 185224 MW; DB3FB2C91F871B CRC64;
 SQ
 Query Match 82.5%; Score 7433; DB 1; Length 1691;
 Best Local Similarity 81.0%; Pred. No. 0;
 Matches 1384; Conservative 121; Mismatches 166; Indels 38; Gaps 5;

Db 370 TAVITAAVYITICHORYLRTQATISKGRRLLEHAOKFISRLYSWLFKSGRDYIPGRQLO 429
 QY 421 FYAQCRRMLASGFLHDPRLVFEDESVPCRCRTFLKVKVAKPCCFMRLGOECTCFLEPAE 480
 Db 430 FYAQCRRMLASGFLHDPRLVFEDESVPCRCRTIRIRIAKPCCFMKWLGQESCFLOPAE 489
 QY 481 GLVGDHGDNEAYEGSEVDPAEPAILDVSQTYAVHGHOLEALYRALNPQDIAARASRLT 540
 Db 490 GLAGDQGHNDNEAYEGSDVTPAEPATLIDIGSYIVGSRSLQTYVALDLPADVARARLS 549
 QY 541 ATVELVASPDRLCECRVTGKNTKRTTYVDCALHEANGPROYVLPDASQSGAGSHSLT 600
 Db 550 ATVTYETISGRDCCQTMIGNKTFLETVDCARLEVENPQOLMSEDSQCSMAAGFCLT 609
 QY 601 YELTLAGLQVRISNGLDCTATFPFGASAPAGEVAACFCSALYRNRTFORHSLTGLW 660
 Db 610 YAAVQGLEVHSTAGLESRYVFPFGNATPAPSEVTACSLALYRNRSQSVIGSLM 669
 QY 661 LHPEGLLGIFPPSPGCHTWESANPCGEGTLYTRWST-----SGFSSDPSPE 709
 Db 670 LHPEGLLGIFPPSPGCHTWESANPCGEGTLYTRWSTITDPLTGLISG-HLDAAPHS 728
 QY 710 AAAPAAATPGLPHSTPPVSDIWLVPPESEEOVAAPVPAADPAGLGPVLPPEPP 769
 Db 729 GCPPTATGPAGVSSDSPD--PLPDVIGSRPSGAR-PAGPNPGV----- 774
 QY 770 PVHKSIPPPSRNRRLLYTPDCAKVVAGSLFESDCDMLVNASNPGHREGGLCHAFYOR 829
 Db 775 -----QRLHLYHPDQAKLYGSIIFESSECTWLVNASNGHREGGGLCHAFYOR 822
 QY 830 FPEATPTFEFIMREGLAATLTPRPIIHAVADYRERQONPKRLEAAYRETCSRGRTAAP 889
 Db 823 YPDSFPAATFVNRDGLAATLTPRPIIHAVADYRERQONPKRLEAAYRETCSRGRTAAP 882
 QY 890 LIGSGIYQVPSLSPDAMERNRHPGDELILTPPAAMVPAAPVLPVLTEDTARTANL 949
 Db 883 LIGAGIYQVPSLSPDAMERNRHPGDELILTPPAAMVPAAPVLPVLTEDTARTANL 942
 QY 950 ALEIDATFVGRACAGCTTSPGIVHGYQFAGYVGSGRKSISQGDVDDVVVVPTRRLNSM 1009
 Db 943 ALEIDSGSEVGRACAGCKVPCGVVRQFAGYVGSGRKSISQGDVDDVVVVPTRRLNSM 1002
 QY 1010 RRRGFATPTPTAARTYIGRRVVIDEAPSLPHLLILHMQRASSVHLGDPNOIPAIDE 1069
 Db 1003 RRRGFATPTPTAARTYIGRRVVIDEAPSLPHLLILHMQRASSVHLGDPNOIPAIDE 1062
 QY 1070 HNGVPAITPELAPTSWNVTHRCRADVCELRGATPKIOTTSRYVRSLFVNNRPAIGQL 1129
 Db 1063 HNGVPAITPELAPTSWNVTHRCRADVCELRGATPKIOTTSRYVRSLFVNNRPAIGQL 1122
 QY 1130 VYTOAAKAAKPAITVNEHOGATFTETIATADARGLIOSSRAHAIVALTFRTEKCVTL 1189
 Db 1123 VYTOAAKAAKPAITVNEHOGATFTETIATADARGLIOSSRAHAIVALTFRTEKCVTL 1182
 QY 1190 DAPGLIREVIGSDVIVNNEFLAGGEVGHHPSPVIRPGNDONLGLQAFPPSCQISAVHQ 1249
 Db 1183 DAPGLIREVIGSDVIVNNEFLAGGEVGHHPSPVIRPGNDONLGLQAFPPSCQISAVHQ 1242
 QY 1250 LAEBELGRHAPPAAYVLPCEPELQGLLYMPQELIYSDSVLVEELDIYVCRMAASQRA 1309
 Db 1243 LAEBELGRHAPPAAYVLPCEPELQGLLYMPQELIYSDSVLVEELDIYVCRMAASQRA 1302
 QY 1310 VLSITLVGRYGRRTKLYEAHSDVRESLAFITPLIGVQATCQELVELYVAMKEGQDGA 1369
 Db 1303 VLSITLVGRYGRRTKLYEAHSDVRESLAFITPLIGVQATCQELVELYVAMKEGQDGA 1362
 QY 1370 VLELDICNDRVSRTFFQCKNKFTTGRTIAGKVGQGISAMSKFCALFGMPRAIEKE 1429
 Db 1363 VLELDICNDRVSRTFFQCKNKFTTGRTIAGKVGQGISAMSKFCALFGMPRAIEKE 1422
 QY 1430 ITALLPNTFYGDAYEESFAAASVAGAGSCMYFENDFSFSDTONNFSIGLECYVMEBCG 1489
 Db 1423 ITLLPQAVFYGDAYEESFAAASVAGAGSCMYFENDFSFSDTONNFSIGLECYVMEBCG 1482

QY 1490 MPQWLLRLVHLVNSAMLLIAPKESLKGFWKKHSGERGCTLLMNTVMNMAIIAHGEFRDR 1549
DB 1483 MPQWLLRLVHLVNSAMLLIAPKESLKGFWKKHSGERGCTLLMNTVMNMAIIAHGEFRDR 1542
QY 1550 VAAFKGDDSVYLCSDYRQSRNAALJAGGGLKLVDPYRIGLAVGVVAPGLGLDLYVR 1609
DB 1543 VAAFKGDDSVYLCSEYRQSRGAGSLIAGGGLKLVDPYRIGLAVGVVAPGLGLDLYVR 1602
QY 1610 FAGRLSEKMWGPGEPEAEOLRLAVCDPLRGLTNVACVQDVYRSYVGVSPGLVHNLIGML 1669
DB 1603 FAGRLSEKMWGPGEPEAEOLRLAVCDPLRGLTNVACVQDVYRSYVGVSPGLVHNLIGML 1662
QY 1670 QTIADGKAHFTETIKFVLDLNTSIORVE 1698
DB 1663 QTIADGKAHFTESVKPILDLTHSIMHSE 1691

RESULT 4
POLN_HEVMT STANDARD: PRT; 1693 AA.
AC 004610;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE
(EC 2.7.7.48); HELICASE].
OS Hepatitis E virus (strain Myanmar) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxId=31769;
RN [1]
RX MEDLINE=9322753; PubMed=8470371;
RA Aye T.T., Uchida T., Ma M.Z., Iida F., Shikata T., Ichikawa M.,
RA Rikihisa T., Wain K.;
RT "Sequence and gene structure of the hepatitis E virus isolated from
Myanmar";
RL Virus Genes 7:95-109(1993).
CC -I- MISCELLANEOUS; HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib.ch).
CC -----
DR EMBL: D10330; BAA01172.1; -
DR MEROPS: C41.001; -
DR InterPro: IPR002589; DUF27.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR000606; Viral_helicesel.
DR Pfam: PF01661; DUF27: 1.
DR Pfam: PF01443; Viral_helicesel: 1.
DR Pfam: PF01660; V_methyltransf: 1.
DR SMART: SM00506; Altp: 1.
KW Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
KW ATP-binding.
FT NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185215 MW; AAB4C9140N7E21EA CRC64;

Query Match 82.5%; Score 7433; DB 1; Length 1693;
Best Local Similarity 81.1%; Pred. No. 0;
Matches 1394; Conservative 109; Mismatches 159; Indels 56; Gaps 6;

QY 2 GITTAEQAALAAANSALANAVVVRPLSLRVGTETLLINMQPQOLYFRRPVLMNPIQRY 61
DB 11 GITTAEQAALAAANSALANAVVVRPLSLRVGTETLLINMQPQOLYFRRPVLMNPIQRY 70
QY 62 INHELEQYCRARGRCLEVGAPHSRINDPNVNLHRCFLRPVGRDVGKWSAPTRGPAANC 121

DB 71 INHELEQYCRARGRCLEVGAPHSRINDPNVNLHRCFLRPVGRDVGKWSAPTRGPAANC 130
QY 122 RSRLRGRLPADRYTCDFGFSRCAFAETGVALSYLHDLMPADYABAMARHGXTLYAAL 181
DB 131 RSRLRGRLPADRYTCDFGFSRCAFAETGVALSYLHDLMPADYABAMARHGXTLYAAL 190
QY 182 HLPPEVLLPPEGYHTTSTYLLIHGDRAVVTYEDTTSAGYHNDVSIIRAWRTTKIYGDHP 241
DB 191 HLPPEVLLPPEGYHTTSTYLLIHGDRAVVTYEDTTSAGYHNDVSIIRAWRTTKIYGDHP 250
QY 242 LVTERVRAIGCHVULLLTAAPESPMPYVPYRSTEVYVNSITPGPGSPLEPSTACTS 301
DB 251 LVTERVRAIGCHVULLLTAAPESPMPYVPYRSTEVYVNSITPGPGSPLEPSTACTS 310
QY 302 TFEHAVVHIMDRMLMFGATLDQAFCSCRLMTYLRGISYKVTYGAALVANEGWASSEDALT 361
DB 311 TFEHAVVHIMDRMLMFGATLDQAFCSCRLMTYLRGISYKVTYGAALVANEGWASSEDALT 370
QY 362 AXITTAAYLTICHOYRRTQALISKMRRLGYEHAQKFTTRLYSWLFKSGRDYIPGQLOP 421
DB 371 AXITTAAYLTICHOYRRTQALISKMRRLGYEHAQKFTTRLYSWLFKSGRDYIPGQLOP 430
QY 422 YAOGRWLSAGFHLDPRLVYFDESVCRCRTFLKVAAGKCCPMRWLGOCTCFLEPABG 481
DB 431 YAOGRWLSAGFHLDPRLVYFDESVCRCRTFLKVAAGKCCPMRWLGOCTCFLEPABG 490
QY 482 LVGDHGDNEAAYGESEVDPAPPAHLVDGTYAAGHOLEALYALVLPDIAARASRLTA 541
DB 491 LVGDHGDNEAAYGESEVDPAPPAHLVDGTYAAGHOLEALYALVLPDIAARASRLTA 550
QY 542 TVELVASPDLRECRYLVGNKTFRTTYVDGAAHLEBANGEOYVLSFDSRQSMGASHSLY 601
DB 551 TVELVASPDLRECRYLVGNKTFRTTYVDGAAHLEBANGEOYVLSFDSRQSMGASHSLY 610
QY 602 ELTPRAGLOVRISSGDCSTAPPPGAGAPSAPEVAFCALRYRFTGRHSITGLTML 661
DB 611 AASAGLEVRVYVAAGDHRVAVFAPGVSPRABEVYAFCAALRYRFTGRHSITGLTML 670
QY 662 HPGGLIGFPFSPGHWESANPFCGEGTLTYRTWS-TSGFSSDFSP--PEAARAPAM 715
DB 671 HPGGLIGFPFSPGHWESANPFCGEGTLTYRTWS-TSGFSSDFSP--PEAARAPAM 720
QY 716 AATPGPLPHSTPPSDIWLVPPESEFQVDAAPVPPAPDPAGLGPVVLTPPPPPYHKPS 775
DB 731 AATPGPLPHSTPPSDIWLVPPESEFQVDAAPVPPAPDPAGLGPVVLTPPPPPYHKPS 785
QY 776 IPPP-----SNRRLTYTPDGAAYVAGSLFESDCMLVNASNPGHPPGG 820
DB 786 IPPP-----SNRRLTYTPDGAAYVAGSLFESDCMLVNASNPGHPPGG 835
QY 821 GLCHAEVYQREPEAFYEFTEIMREGLAAYTLTPPPIIHAAVADYRVQDNPRLDAAYRETC 880
DB 831 GLCHAEVYQREPEAFYEFTEIMREGLAAYTLTPPPIIHAAVADYRVQDNPRLDAAYRETC 885
QY 816 GLCHAEVYQREPEAFYEFTEIMREGLAAYTLTPPPIIHAAVADYRVQDNPRLDAAYRETC 875
DB 875 GLCHAEVYQREPEAFYEFTEIMREGLAAYTLTPPPIIHAAVADYRVQDNPRLDAAYRETC 880
QY 881 SRGTAAYPLGSGTYQVPSLSPDAMERNHNRGDELYLTERPANNPEAKRQPVLTIT 940
DB 891 SRGTAAYPLGSGTYQVPSLSPDAMERNHNRGDELYLTERPANNPEAKRQPVLTIT 945
QY 941 EDTPARTANLALIEDATVEVCRACAGCTISPGIYHGYOTAGVPSSGSRSTTQGDVAVVY 1000
DB 951 EDTPARTANLALIEDATVEVCRACAGCTISPGIYHGYOTAGVPSSGSRSTTQGDVAVVY 1005
QY 936 EDVARTANLALIEDATVEVCRACAGCTISPGIYHGYOTAGVPSSGSRSTTQGDVAVVY 995
DB 996 EDVARTANLALIEDATVEVCRACAGCTISPGIYHGYOTAGVPSSGSRSTTQGDVAVVY 1000
QY 1001 PTERLNSMRRGFAAFETPHTAAYTIGRRVYIDEAPSLPPLLHLMORASSVHLGDP 1060
DB 1011 PTERLNSMRRGFAAFETPHTAAYTIGRRVYIDEAPSLPPLLHLMORASSVHLGDP 1065
QY 996 PTERLNSMRRGFAAFETPHTAAYTIGRRVYIDEAPSLPPLLHLMORASSVHLGDP 1055
DB 1055 PTERLNSMRRGFAAFETPHTAAYTIGRRVYIDEAPSLPPLLHLMORASSVHLGDP 1060
QY 1061 NQIPADDEFHAGVYPAIRBELAPTSMWXYTHRCRADVCEILKATYKIQTTSVYLSLFW 1120
DB 1071 NQIPADDEFHAGVYPAIRBELAPTSMWXYTHRCRADVCEILKATYKIQTTSVYLSLFW 1125
QY 1056 NQIPADDEFHAGVYPAIRBELAPTSMWXYTHRCRADVCEILKATYKIQTTSVYLSLFW 1115
DB 1115 NQIPADDEFHAGVYPAIRBELAPTSMWXYTHRCRADVCEILKATYKIQTTSVYLSLFW 1120
QY 1121 NEPAIGOKLYXQAAKAAANGALITVEAGATFETTTIATADARGLISSRAHALVAL 1180
DB 1181 NEPAIGOKLYXQAAKAAANGALITVEAGATFETTTIATADARGLISSRAHALVAL 1185

```

Db 1116 GEPANGQKLVFTQAKANPGSVTYHEAGATYTTTATADARGLIOSSRAHAIVALT 1175
QY 1181 RHEKCVILADAPGLREVGISDPVYNNFVLAGEVGHSPVYIPRGNPQNLCTLOAFP 1240
Db 1176 RHREKCVILADAPGLREVGISDAIVNNFVLAGEIGHQSPVYIPRGNPANDTLAAPP 1235
QY 1241 SCQISAYHOLAEELGHRPAVAALPPCELEGLLYMPQELTVSDSVLVELFDIVCR 1300
Db 1236 SCQISAFHOLAEELGHRPAVAALPPCELEGLLYLPQGLTACDSVTFELTDIVCR 1295
QY 1301 MAAPSRKAVLSTVIGRYRRKLYNAASHDVDSLAFIPALGPVOTTCLEYELVEM 1360
Db 1296 MAAPNRKAVLSTVIGRYRRKLYNAASHDVDSLAFIPALGPVOTTCLEYELVEM 1355
QY 1361 VERGOGSAVLELDLCNRVSRITTFQKCKNFTTGETTAHGKVGOGISAMSKTFCALG 1420
Db 1356 VERGOGSAVLELDLCNRVSRITTFQKCKNFTTGETTAHGKVGOGISAMSKTFCALG 1415
QY 1421 PWFRALEKILALPPNIFYGDAYEESVFAAASGAGSCWPEENDESEFSTQNNFSLGL 1480
Db 1416 PWFRALEKILALPPQGVFYGDADFTVFSAAVAAKASWPEENDESEFSTQNNFSLGL 1475
QY 1481 ECVYMECECMQWMLRLYLHLVSAWTLQAPKESLKGFMKKHSGEPGLTMMYNNAAITA 1540
Db 1476 ECAIMECECMQWMLRLYLHLVSAWTLQAPKESLKGFMKKHSGEPGLTMMYNNAAITA 1535
QY 1541 HCYEERDFRVAAFKGGDSVVLCSYQSRNAALVAGCLKLVDPYRPIGLVAGVVAAG 1600
Db 1536 HCYEERDFRVAAFKGGDSVVLCSYQSRNAALVAGCLKLVDPYRPIGLVAGVVAAG 1595
QY 1601 LGTLPDVVFAAGRLSEKNNKNGPGERAEQLRLAVCDELRLTNVAQVCDVVSRYGVSPG 1660
Db 1596 LGALPDVVFAAGRLSEKNNKNGPGERAEQLRLAVSDELRLTNVAQVCDVVSRYGVSPG 1655
QY 1661 LVHNLIGMLQIADGKAHFTETIKPVLDTNSTIQRVE 1698
Db 1656 LVHNLIGMLQAVADGKAHFTESYKPVLDLTNSTICRVE 1693

RESULT 5
RPRO_CGMWS STANDARD; PRT; 1648 AA.
ID RPRO_CGMWS
AC P19523; Q83208; P89877; P90356;
DT 01-FEB-1991 (Rel. 17, Created)
DE 20-AUG-2001 (Rel. 40, Last sequence update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (129 KDA PROTEIN)].
OS Cucumbe green mottle mosaic virus (watermelon strain SH) (CGMW), and
OS Cucumbe green mottle mosaic virus (watermelon strain W) (CGMW).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=12236, 12237;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SH;
RX MEDLINE=91311400; PubMed=1856687;
RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
RA Sato T., Motoyoshi F., Nishiguchi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 1496-1648 FROM N.A.
RC STRAIN-W;
RX MEDLINE=89073773; PubMed=3201760;
RA Sato T., Imai Y., Meshi T., Okada Y.;
RL "Interviral homologues of the 30k proteins of tobamoviruses.";

```

```

RL Virology 167:653-656(1988).
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: REAPPROACH OF THE TERMINATOR UAG OCCURS BETWEEN
CC CODONS FOR LYS-1144 AND GLN-1146.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D12505; BAA18895.1; -
DR EMBL: D12505; BAA18896.1; -
DR EMBL: J04322; AAA46382.1; -
DR PIR: J01157; WMTMS2.
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR006066; Viral_helicase.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
KW Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1648 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 863 870 ATP (POTENTIAL).
SQ SEQUENCE 1648 AA; 186549 MW; 1D1AFEEB/65595 CRC64;

Query Match 4.5%; Score 406.5; DB 1; Length 1648;
Best Local Similarity 18.8%; Pred. No. 4.3e-18;
Matches 350; Conservative 220; Mismatches 633; Indels 661; Gaps 73;

QY 77 CLEVA-HPRSINDNPVNLHRC-----FLRPVGRDVQRMYSAPTRGPANCRSALRGIP- 130
Db 105 CYDIGNTQYHLFKGRSYVHCNPNCLDLKDVARNMV-YNDMTITQYQRIKSGSGCRPLPT 163
QY 131 ---PADRY-----CFDGSRCFAFAETG-----VALYSLHIDLPADVAEMARRGX 174
Db 164 PQIDFRRYDSSPCAVTSCSDVFOECSYDPSGSDRDHNAVLSHSIYDIPYSIGPALRRKV 223
QY 175 TRLVAAHLHPREVLL--PRGYHTSYLLIHGDRADVITYBEDTSAGYHHDSILRAWR 232
Db 224 RVCYAAFHSEALLLGSPVGNLNSIGAQFRVGDVDVHFLFSESTLATHSLENIKLYM 283
QY 233 TTKIVGDHPVLER--VRAIGCHFVLLLTAAPEPSMPYVYPRSTEVYVRSIFGPGS 289
Db 284 RYTFPADRDFVYIKEFMKVRVDYTFFFRLVRADTIHLHKSVMGYSKSKSEYFL-----NT 338
QY 290 PSLFSPSCKSTGTF-----HAYVNIW 311
Db 339 PPIF-----QDKATFSGWPEAKRKVLIPKFEISREFLSGNVKISRMVADAFVHTIINH- 393
QY 312 DKLMLFGATLDQAFCCSLMTYLLIGISIKYVVGALVNAEGNNAEDLTAITAYLLI 371
Db 394 -----STYDNKALVWKNVQSFVESIRSRVYVNGSVKSENNVPVDDLTIDISFIFLV 446
QY 372 CHQRYLTQAIKSGMRRLGVEHAQKFTIRLYSWLFEKSGRDYIPGROLFVYACRRWL 431
Db 447 -----KYRKQIE----- 454
QY 432 GFHLDPRLVFEDESVPCRCRFLKRVAGKFCCEMRLQOECTCLEPAEGLVDGHGHDNE 491
Db 455 -----LMSDKVYIEARGLLRFADELKSAVGLG-DCV-----YDALVGTGWFPTS 499
QY 492 AYESGEVDAPEPAHLDVSTYAVHGHOLEALRYALNVPQDIAARSLRTATVELVASPDR 551
Db 500 SDELVLVLP-EP-----FWTFPSDYLEGVYEA-----DAKIERESVS-ELLASGSD 542

```

QY	552	L	-	ECRFLVGNKIFR	-----	TTVVDAHLEANGPEQYVLSPDASRQSMG	593			
Db	543	L	FKKID	IEDIRNNYSGVGEFDEVEKFOEFCCKELNVNPMGLIGHVIE	-----	IFSOKAGVYTG	596			
QY	594	AGSHLSLEYLTP	-	AGLOVRRSSNGLD	-C-----	TATPPGAGSAAPGEVAACS	641			
Db	597	LGT	-----	LSPEKGSVALSTSVTCCEDMDYTEDMEDIVIMADKSHSYSPENAKRAD	-----		650			
QY	642	ALYKRNFTORHSLTGGMLHMLPEGLIGIPPPSPGHIMESANPCGEGTLYRTWSTSGF	701							
Db	651	VKYGNMK	-----	GALVEYKVGTSMTLP	-----	ATNAEKG	679			
QY	702	SSDSEPEEAAAPAMATPGLPHSTPPVSDIWLPP	-----	PSSE	-----	FQVDA	746			
Db	680	-----	-----	KAVLPSGICVAKPKQFSKPLEDEDLRLSNNEFKVSDL	-----		718			
QY	747	PVPAPDPAGLPGPVLTTPPPPVHAKPSLPPSRNRLLTYPPGAKVYGS	-----				799			
Db	719	KLKKTITPVYTGTT	-----	REKQKMYIDYLSASLGSTGNLER	-----		758			
QY	800	LFESD	-----	CDWLVSANPGHRPGGGLCHAFYOREPEAFYPTFEIM	-----		841			
Db	759	IYSDMNGTEESMQTGLYDCEKCKAKMLLP	-----	PAEKHAMAVYLA	-----	SDJT	805			
QY	842	REGIAATYTLPRPIIAHAVDYVEQONPKRLEAAVRETCSRGTAAVPLLSGITYOVPS	901							
Db	806	RIIFLSYDESGSLID	-----	KKMKRF	-----	TKVSVIRS	842			
QY	902	LSFPAMERNRPODELYPEAPNMEAKPAQPVLTITEDARTANLALIEDATVEGR	961							
Db	843	LEV	-----	LKNEA	-----		850			
QY	962	ACAGCTISPGIVHYOFTAGVPGSGKSRSI	-----	QOGDNDVVVPTRELNSMRBRGAFT	-----		1018			
Db	851	-----	YDPG	VHITLVYDVGPGCGKAEIILARYNMKMTDLVLPGRGAAMIRRAALHK	-----		904			
QY	1019	PHYAR	-----	VTIGRR	-----	VIDEAPSLPHLL	-----	LHKORASSVHLGDPNQ	1066	
Db	905	SPVATNDNVRTFDFSEVNNRKIFEFKADVAYDEGLMHTGTLNFALKISGCKAKFAVYDGAQ	964							
QY	1063	IPAI	-----	DREHGLVPAIRPELAPTSMMXVTHRCADVCELIRGAY	-----	PKIOTSTRVLS	-----		1111	
Db	965	IPFTNRWMDYPRKELTTLVDNVERRY	-----	VTHRCPRDYTSFLNTYKAAVATTSVPHS	-----				1022	
QY	1118	LFENNEPAT	-----	GOKLYXTQA	-----	AKANPGAIYTHEAQATFEY	-----		1156	
Db	1203	V	-----	KAIKYSAGILARBELTKIKGKITFTQSDKOSLKSQNDVNYVTHEIQUETFEET	-----				1076	
QY	1157	TI	-----	ATAADAGLIOSRAHAIVALTRTEK	-----	CVILDAPGLLREVGISDYVNNFLA	-----		1211	
Db	1079	AVVATPTPIGLIARDSPHYVALTRHTKAMYVYTVAFDAVTSI	-----	IADY	-----				1126	
QY	1212	GGEVGNRPVSITPRGNDQWLGLQAFPPSCQISATHQLAEBLGNRPAAVAVLPRCEL	1277							
Db	1127	-----	-----	EKYDOSILTM	-----	FATVPTKXOL	-----		1147	
QY	1272	EOGLIYVPOELTVSDVLYELNDIYHCHMAAPSOKAVLSTLYGRGR	-----						1326	
Db	1148	MONSLY	-----	VHRNIFLVSXGTGYTDMQEYTRDKLCPGNSFVLD	-----				1188	
QY	1321	-----	-----	RTKLEYAANDVRESLARFIPT	-----	IGPVQATTC	-----	LYELV	1357	
Db	1189	FDAVTMLRLRNEFMLOPCRLTILSNLDPVPALIKNEAONFLIPVLRYACERPRIPGLEML	1248							
QY	1358	EAMVEKQDGSAYL	-----	EELDCNRDVS	-----	RTTFQKKCKNFT	-----		1396	
Db	1249	VAMIKRMMNTPDLAGTVDITNMSISIVDNFESSFVDEVLDLHDCVRASSIOSKSWFS	1308							
QY	1395	TGETIARGKVG	-----						OGISMSKTFQAL	1418
Db	1309	CQPSANVGOLANRFIDLPAFTYIMIMIKRQKRSJLDTISQSEYPALQITVYHPRVYNAV	1366							
Y	1419	FGPFRAIERKEILLALPPN	-----	IFYGDAYESV	-----	FAAAVSAGSGCVAFENDSEEDSTON	-----		1476	

Db	1369	FGVFFKLTTRKFLSMWDSKFFFTYTRKKREDDLOEFFSDLSHSDYEILFELDYSKYDSQS	1428
QY	1475	NFSLGLECVYMECGMPQWILIRYLHVRSAMILQAPRESLKG--FWKXHGSEPTGLMT	1532
Db	1429	DFHESIMATWELKGLDDLLAMWMSNGHRTLLDFOAGIKTLIYORKSGDVFPTIGNT	1488
QY	1533	-----WNNAAITAHCEYEFDFEVAAFKCRDDSVLCS---DYRQSRNNAALITACGLTKLK	1583
Db	1489	FIINACVASHLPIDKCK-----AFCGDDSLIYLPKGLERPDQIANTLV--WNFEAK	1540
QY	1584	VDYRPGILLYGVVY---APQLGLTPDVYRPAGRLSKKNMGPEPEAEQDLAVCDLFGRL	1640
Db	1541	LERRKYVPECGKYIIHHANCVIYPPPLKLISKLNKSL-VGEYEHVEEFRSLDVAHSL	1599
QY	1641	TNVA	1644
Db	1600	FNCA	1603

```

RESULT      6
POLN_RUBVYT ID POLN_RUBVYT STANDARD: PRT: 2205 AA.
AC          13889;
AD          01-JAN-1990 (Rel. 13, Created)
DT          01-AUG-1992 (Rel. 23, Last sequence update)
DT          20-AUG-2001 (Rel. 40, Last annotation update)
DE          NONSTRUCTURAL POLYPROTEIN [CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO
DE          NSP4].
OS          Rubella virus (strain Therien).
OC          Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC          Rubivirus.
OX          NCBI_TaxID=11045;
RN          [1]
RP          SEQUENCE FROM N.A.
RA          MEDLINE=90281585; PubMed=2353453;
RX          Dominguez G., Wang C.Y., Frey T.K.;
RT          "Sequence of the genome RNA of rubella virus: evidence for genetic
RT          rearrangement during togavirus evolution.";
RL          Virology 177:225-258(1990).
RN          [2]
RP          SEQUENCE OF 1737-2205 FROM N.A.
RA          MEDLINE=88226020; PubMed=2836271;
RX          Frey T.K., Marr L.D.;
RT          "Sequence of the region coding for virion proteins C and E2 and the
RT          carboxy terminus of the nonstructural proteins of rubella virus:
RT          comparison with alphaviruses.";
RL          Gene 62:85-99(1988).
CC          -I- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC          -----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration
CC          between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC          the European Bioinformatics Institute. There are no restrictions on ways
CC          use by non-profit institutions as long as its content is in no way
CC          modified and this statement is not removed. Usage by and for commercial
CC          entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC          or send an email to license@isb-sib.ch).
CC          -----
DR          EMBL, M15240; AAA88528.1; -.
DR          PIR, A35320; MNWVRN.
DR          MEROPS; C27.001; -.
DR          InterPro; IPR002589; DUF27.
DR          Pfam; PF01661; DUF27; 1.
DR          SMART; SM00506; ALPP; 1.
RW          Polyprotein; Nonstructural protein.
FT          CHAIN          1          ?          NONSTRUCTURAL PROTEIN NSP1.
FT          CHAIN          ?          ?          NONSTRUCTURAL PROTEIN NSP2.
FT          CHAIN          ?          ?          NONSTRUCTURAL PROTEIN NSP3.
FT          CHAIN          ?          2205          NONSTRUCTURAL PROTEIN NSP4.
SO          SEQUENCE          2205 AA; 240220 MW; 15A95F5E534C0B03 CRC64;

```

Best Local Similarity 19.2%; Pred. No. 7.5e-16;
Matches 454; Conservative 201; Mismatches 714; Indels 995; Gaps 104;

QY	3	ITTAEOAAALNAANSALANAVVUPVPSRQOTELLINLMQROQLVPREVILMNPIORVI	62
Db	40	VYTAOKKAIY-----AIVPRVFTOMYS-----DHP---AL	69
QY	63	HNELEOYCRARAGCLEVGANHR-----INDNPNVLHCRFLRVGNDVOMYS-----	111
Db	70	H-AISRYTRR---HMIENG--PKAHLVILIPSPGL-----LREVARERVRWALCLHRT	118
QY	112	APTRGRAN-----CRSALRGILRP-----ADRTIC-EDGF	144
Db	119	ARKATATALETAASEKMHADYVC-----ALRGASGPRFYVNHEDVPRGGAUADRCLLYTPM	175
QY	142	SRCAFAETGYALYSLNHIMPADYA-----E	167
Db	176	OMCELMRTIDATLLVAVDLMVVALAANHGDDMDLGIAMHLHDGGCRADCRGAGAPTR	235
QY	168	AMARHGKRLYAALHPRVULIPRTYITTYLLIHGDRAVVUTYEGDTSAGYHHVYSIL	222
Db	236	GUTRCSTRIVQVL---PDTAHPGLYKRGPRILTRDCAVALLESWEVAQHCGR-----	286
QY	228	RAMITTKIVGDHPLVIERVAIGCHFYLTLTAAPERSPMRYUVPRSTEVYVSIFRG	287
Db	287	-----ARVAIVKCTLPIRNVSLQPS-----ARV-----	311
QY	288	GSPSLFPASGCTKSTENAVPHVINDRLMLFEGATLDDOAFCCSRIMTYLRGISYVTVGAL	347
Db	312	LPDLVLAEGVGRWFWFSLPVPOVMLSYCKTSLSPRAYSERFVKKNALCHSTILAGN	370
QY	348	VANEGMN---ASEDLTXHTAATVLTICHQVLFRTQALSKGRRLGVEHAKCFITRLYS-	403
Db	371	VLOGWKKCTCAEBDALCA-----YVAFRAMQSANRAGIMKKGAKCAADLSV	411
QY	404	-WLEKSGRDYIPROLOQFYAOCRWLSA-----GFHLDRVLVFDSEVBCRCRTE	455
Db	418	AGWL-----DTI-----WDAIKKRFGISVPLAERMEWEQDAVAAFD-----	454
QY	454	LKKVAGKFCSCMRYMLGQECSTGLBPABEVLVDHGDHNDNAEYGSVDAAERHILDSGTYA	514
Db	455	-----KQLEDDGGRHLDYVO-----PKSPRPRETAITWI	484
QY	514	VHGHOLE---ALYRALNVPD-----IAARAS	537
Db	485	VHAASEDNHCACARCDVPRERPARAGORDDALLIPWLEAERRALRCREMDFEALRAR	544
QY	538	RUTATVELVASPDLECTVL-----GNKFRFTYVDGALLLEANGEOY	584
Db	545	ADTAAAPRPARPRARKYTYLYKRPANHGRPVLTIDREBEADALVLCRQGLRGERH	604
QY	582	VLSFDASQSGAGSHSLTYELTRAGLO--VRISNGLDCTATPRPGASAPARGEVAAF	638
Db	605	-----FAAGAHMCA---QARGLOAFVRY-----VPRPERPADGG--ARA	639
QY	640	CSALYRYNRFTQORHSLNGMLNHRGELG---IFPPSPRHIMESANPRCGESTLYTRT	699
Db	640	WAKFFRGCAKOR-----LIGERAVMHLPTDGDVPOLI-----ALALRT	679
QY	696	WSTGFSFSDFS---PREAARPAMAAT-----GGLP-----HS	724
Db	680	LAQOGAALALSVRDLPGGAARFADANNAVTAAVACGRQSAASAPRGDPRPRARARSRHS	739
QY	725	TPPVSDIWLVRPSEEFQVDAAPVPRAPRDAGLRGBVULTRPPRRPNK-----	773
Db	740	DARGTPRPARADPPR-----APSPRAPRAGDVP---PIAGRADRARDAELE	787
QY	774	PSIPPSRNR-----RLTYTPDCAKUYAGSLFESDCWL	808
Db	788	VACERPSGPRITRADPDSDIESTAKARAAGPHLAVRDIIMDRPCKV-----V	833
QY	809	VNASNPGRHGGGICHAFYORFPFAFYPTFEIMREGIAA--YTLTPRP-----	854

Dd	836	VMAANEGLLAGSGVCGAIFANATATA-----	LAANCRLRAPCPTEGEAVATPCHGC	884
Qy	835	-----IHAAVP-----DYREONPKLEAAYRE-----	TCSRGRTAAPPLAGSGIY--QV	898
Dd	885	GUTHIIHAIVAVRRRDBPAALIEEGEALLERAYRSIVALAARMAACVACPLLAGAGYNSA	944	
Qy	899	PVSLSEFDAMEENHRPGDELY-----	TEPANMFE	928
Dd	945	AESLRALAAARTPEPVERVSLIHCHPDRALTHIASVLAGLAARVSPPTPEPLAS-CP	1003	
Qy	929	ANKPAAPVLTITTEDTARTANILALEIDAAYEGHACAGCTIS-----	969	
Dd	1004	AGDEGRPAORSASPATPLG-----DATAPEPBGCGCELCRYTETVNDRAVYNLMTERPD	1058	
Qy	970	-----PGIY-----	HQF-----	977
Dd	1059	RGATSMAMRIPEVVVYVPEHLATHPLNHTSVLKRAEVRPPRGMGCSDDMRCHGHCMPQ	1118	
Qy	978	-----TAGVPSGSKSRNSIOCGDVVVVPTRELRN-SMRRGFAAFTPTAA	10233	
Dd	1119	VRCPPSNMAHALCRTGVPPASTRG---GELDPNCTWLBAAANVAQAAACGAYTSACCP	11757	
Qy	1024	RVTIGR-----	RVVIDAP	1037
Dd	1176	KCAVGRLSEARTHEDEFAALSQKMSASHADASPDGTGDLPLMETVGCACSRVWVGSEH	12355	
Qy	1038	SLPRLHLLHQRA-----	SSVHL-----GDPNOI--PAI-----	1066
Dd	1236	EAPDPHLIVSLHRPNCPMGVLEVARAPBGCPNTHFCANAGGGRNRSDPLHMLAVP	12955	
Qy	1067	-----DFE-----	HAGLVPAIRPELAPTS--WMXYT---	1090
Dd	1296	LSRGGTCATDEGLAQAYVDDLEVRRLGDAMARAALASVORPKRGYNIRVWMMAGA	13555	
Qy	1091	-----HGPCA-----DVCELR-----	GAYPKI--	1108
Dd	1336	GKTRILLATFREDLVYCPTNALLHEIQAKLRADIDIKNAATYERUKLTPLAATRYI	1415	
Qy	1109	-----OTTSRL-----	RSLF-WNEP-	11233
Dd	1416	DEAFTLGGEYCAFAVSQTTAEVICVGDRDCCGPHYANNCRTEVPDPRWPTERSRHTWREPD	14757	
Qy	1124	-----	AIGOKLVXQAAKANPGLIYHE	1147
Dd	1476	CWAARLRAGLDYDIEGERTGTACNIMDGRQVDLHLAFSRETIV--RLHEAGITRYIYRE	15333	
Qy	1148	AQAGFTTETTTIATADARGI--IOSSRHAIVALTRHTECSVILD--APGLREVGISDIYV	1205	
Dd	1534	AQGSVGTACIHVGROSTDVALMLTDLIAVSLTRASDALYIHELDEDSLRAGUS----	1589	
Qy	1206	NNFLAGEV-----GXHR-----	PSYPRGNPNQIGTIQAPRPSQISATYHQL	1250
Dd	1590	-----AFLDGAALAEIKEVPAAGIDRVAAVEQAPPLP---PADGIREADQVPPFCPRLEELV	1644	
Qy	1251	AEEGLHRAPAVAALVPCPELEOGLLYMPOELTVSDSVLVEFLTPIVHCRIAPASORKAV	1310	
Dd	1645	FGRRGH---PHYADLNVTYBGEREYKX-----RISHLLNKHNTDM-----	POTER-V	1689
Qy	1311	LSFLVGRYGRRTKLYEAAH--SDVRESLARFIP---TIGVOAAT-----	CELEYL	1356
Dd	1690	LSAVCA-----VRRYRAGEGSGTLRTAVAROHPRPRQRPPIPRVYAGVAOEWRTMYLBER	1744	
Qy	1357	VEAVWEKGDOSSAVLEL-----	DLGNRVSRITTEFOKXCKMF-----	1394
Dd	1745	IDLTDVYQKVAARBELTDRIYARRPEIIPAGMCTAOSLSVPFLKATIKCYDAALGPMDT	1804	
Qy	1395	TGETIANGKVGOGISANSKTFCCALFGWFALEKELIALLPNIIFYGAYESVFAAAS	1454	
Dd	1805	EDCHAAGKAGLEIRAAKAKWQVMSHPRAIOKIMRLARQPLVAAGHTEPEVDAMQ	1864	
Qy	1455	GAGSCWPFENDESEFDSTONNES-----	IGLECVMEEGCMQMLIRLYHLVRS	1503
Dd	1865	AHYTTNIAEVDPEFMDNQULATRDVELEISALLGLPCA--ED-----	YRALRA	1912

```

OY 1504 -AMILQARKESUJCKGFWKXHSBPGGLLNTNTYMMALIALHACYEER---DFRVA-A-FKGGD 1557
Db 1913 GSYCTPLRELGSJETCCECTSEPALHTNTT-----VAMCMMARVPKGVRAAGIFQGGD 1967
OY 1558 SVLCSIDYROSNNAAAL-----IACCGLKLVDPYRPIGLYAGVVAAPGCTLPDYVR 1609
Db 1968 WYIFLPEGARS---AALKWTPAEVGLFEGFHIPKRVSPPTSPFCG-HYSTAGLEPHDYHH 2023
OY 1610 FAGRLSEKNMGPEPERAQLRLAYCDPLRTTNVAQVCVYVSRYGVSPGLVHNLIGML 1669
Db 2024 QAIKVLGRFR--DPDVEEQVALLDRLRG-----YYALP----- 2057
OY 1670 QTIADGKAHFTETTKPVDLTNSTI 1693
Db 2058 DTVANAAAYDYDSASERVLAIVREL 2081

RESULT 7
RPRO.ORSVS
ID RPRO.ORSVS STANDARD; PRT; 1612 AA.
AC 084133;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (BC 2.7.7.4b) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSPERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Odontoglossum ringspot virus (isolate Singapore 1) (ORSV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138662;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96257213; PubMed=8666266;
RA Chung C.G., Wong S.M., Mahiani P.H., Loh C.S., Goh C.J., Kao M.C.C.,
RA Hong M.C.M., Matanabe Y.;
RT "The complete sequence of a Singapore isolate of odontoglossum
RT ringspot virus and comparison with other tobamoviruses.";
RL Gene 171:155-161(1996).
CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSPERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR LEU-1112 AND GLN-1114.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration of
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; U34586; AAC55012.1; -
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR002588; V_methyltransf.
DR InterPro; IPR000606; Viral_helicasel.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicasel; 1.
DR Pfam; PF01660; Vmethyltransf; 1.
KW Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1612 RNA-DIRECTED RNA POLYMERASE.
FT NP_BIND 1 1112 METHYLTRANSPERASE/RNA HELICASE.
FT FT 826 833 ATP (POTENTIAL).
SQ SEQUENCE 1612 AA; 183194 MW; 3BA44670B0ACB189 CRC64;

```

Query Match	3.9%	Score 351.5;	DB 1;	Length 1612;
Best Local Similarity	19.5%;	Pred. No. 1.5e+14;		
Matches 357;	Conservative 226;	Mismatches 657;	Indels 589;	Gaps 84.
QY	94 LHRCE----	LRPVGADVDVQWTSAPPRGGAANCRRSALGCLP-----	PADRYTC	137
	: :	: :	: :	: :

Dh	123	VHCOMPNIDIDVARHINQDDVSTY--LARLERSK-RGLPVRQGSFAFKMYMNDPAVOC	179
Oy	138	FDGFSRCFAFA-----ETGVALISLHDJMPAD--VAEAMARHGXTRLALHLPREVLEPP	191
Dh	180	DKRRDQSYSSAGLPGKTYAVAGLSIYDI--PADEFGALLKRDHVICALUYAFAHSEENILL--	236
Oy	192	GTYHTSTYLL-----IHGDRAVVTYEDDTSAGNHQVSLIRARINT-----KIMG	238
Dh	237	---ETTSAPDEIGATYTKSGDRLSFFQNESTLNEHSTKYKVIKVCCTPPASNRFPAY	293
Oy	239	DHPLVIEVRAIGCFHILLT-----AAPSPMPYVPYPRSTEVVRSIFGCGSP	290
Dh	294	HKEKMCRTVRNWFCKFPFKVDTYELFRGVYTRGEDSDQFTIADNEANE-YKTLAMINSR	352
Oy	291	SLFSPASCTKSTFHAVP---VHIMDRMLFG-----ATLDQGA	325
Dh	326	FCCSRIMTYLGRISYKXTYGCALVANEEMWNSSEDALTRAXTAALVTICHQRYLRTQASIG	385
Oy	413	LYTKNVLSFEVSIKSRVYINGVTARSEMDVDSYLA--LSMTPL-----LOTK-LABA	463
Dh	386	MRRGLVEHAQF---ITRLYSMLEKSGRDYIPGRQLOFAOGRMLSGFHLDPVYLF	442
Dh	464	KDQVYLLKFKQFEDTYINLFWKQJSDAVGLFSPK-----ERLISGF-----YKVA	511
Oy	443	DESVPSCRFTFLKKVAKGFCCFMWLGQECTCELEPAEGLVGDHGDNEAEGSEYDPAE	502
Dh	512	EQSLQIKTPDE-----YITPADKLME-----YQATE---E	539
Oy	503	PAHLDVSTYVHGHOLEALYRALNODIAARASRLTALVELVASPDRECTVYGNKT	562
Dh	540	LOHLDISKPL-----ERAEKYNYAL-----SELSTYKE--CDEFDITQFKNICEAD	584
Oy	563	FRTYVQDAHLEANGPRQYVLSFDASRQSGAGSHSLTYELFPAQGOVRISNGDCTAT	622
Dh	565	IAPDV-----AKYVFIKMKNELTL-----	604
Oy	623	FPFGGAPSAABEVAALFCSALYRYNRPTQRSHSLTGLMLHBEGLIGFPFSPGHIMESA	682
Dh	605	-----PFK-----	607
Oy	683	NPFQGEGLTYRTMSTGFSFSDSPRPAAPAMAATPG-----LPHSTPRV	728
Dh	608	NP-----TPALSDALSPKPLDLMRFLCLKLSTCQAPFSVKT	645
Oy	729	SDIWLPPPS---EEFYDADAAPVPAPRPAQLPGRVYLTPRRPVYKRPSPRSSRRKLL	786
Dh	646	LDGSLGPKQOSGDRQOSESQSVSVSD-----FLIKSESVKXKSM--	686
Oy	787	YTYPDGAARVYAGSL-----FESDCWL-----VNASNPGHRRGGGLCHAF---YQRPPEAF	834
Dh	667	-----SSAVYVGLPKVOOMKNYMDYLSASISATVSN-----LCKVYLKDYIADPES-	732
Oy	835	YPTFEMREGI-----AAYTLTPRPILHA--VAPDYRVEDONPKRLAEAYRETCSRGTAA	887
Dh	733	-----AEKSGYDVYKGMKMLLKPKDKCHAMGAVELNNGEKVYIULLEMA-----DG	777
Oy	888	YPLGSGIYQVPSLSDANERNHRPDELLYTE--PAANWFPAKRPAPQVYLTIEDTAR	945
Dh	778	FPICGD-----WRYAVSSDSPISYDDMGLQTLLSCLDGDGPVYLRMPKYT--	822
Oy	946	TANLALFIDATEVGRACAGCTISPGIVNHQTFAGVPGSGKSRSIOG--DVDVYVYPT	1000
Dh	823	-----LVDSVPGCGCKTELLEYNFNDEDDLIVNG	851
Oy	1003	RELNSWRRRG-----FAAPTPTTARAVTIGRRVVYDEAPSLRPHILLH	1047
Dh	852	KEACKMIIRKANKSGHVRATRDVRYVJSTLMLKPKTY--NKLFLDEG-----LMLH	902
Oy	1048	-----MORASSVHLCDPNOIPAIDF-----EHAGLVPAIRPELAPTSMWXYTH	1091
Dh	903	TGCVNPLVALSHCREAVYFGDADQIPEFINVAFPYPKFRYRYCLNHRREVRRLST-----	957

QY	776	IPPSRRRLRLLYPPGCAKVASJESPCDWLVNA--SNPCHRRPG-----GLCH--	824
Db	776	LPOPTLN-CLLSAVSDQTKVSEEHMEWSIQTLTPRPSQUSNEBNTTLGLSTHLLTAHLHY	834
QY	825	-----ARYQRPPEAFYPTFERIMREGIAATYTLPRPLIAVAAPDYVEQENPKREAAAYRET	879
Db	835	NFOATVYSDKGPLILFEPESDPIKR-----IDITHGTGRPHSFPG-KRLGSDP--	881
QY	880	CSRGRTAAYVLLG-----SGTYQVPSLFDAMERNHREGDELY---LTEPAANFEA	929
Db	882	-SAGHPSDDLIRAMKSFKSGNRY-LPFS-----EAHNHPTISHAKNLISNNKNGPDG	933
QY	930	NKRAQPLVLTTEBTATANAILEIDATAEVGAACAGCTISPG-----LYHYQFTAGVQSG	985
Db	934	VLSLDDVTSQRTGTPPKERITIQIDHYLYDT-----NPKCTTPVYHF--AGFACG	981
QY	986	KSRSDIOQ-----GVDDVYVYPTRELRSMMR---RCFAA--PTPTTAARVTIGRRVY	1032
Db	982	KTYIYQQLKTKLFKDFR-VSCPTTLEKRTKMTAMELHSGOSWRNTWESSLASSRLY	1040
QY	1033	IDEAPSLPRHLLHLMORASSVHL--LGDV-----NQIPADIEFHAGIYV	1075
Db	1041	IDETIKYMPRGYLDLSILADPALELYITLGDPIQGEYHSGSKDSSNHRIPS---ETLRLLP	1097
QY	1076	AIRBELAPTYMWXVTHRCADVCCELL-----RGANPKIOT-----TSKXLSLF	1119
Db	1098	YI-----DMYCMW--SYRIPCCIRLEQIHSFNAMQSGVIGSVSTPHDQSPVLTNSHASTLT	1151
QY	1120	WNEPAIOGXVTHQAAKANPGAITHEAGCAFTEFTTITATADARGLIOSRAHAVAL	1179
Db	1152	FN--SLGYR-----SCTISSQGLTFCDPAITVLDNTKWLSS--ANGVAL	1194
QY	1180	TR-----HTEKCVILDA--PGLIREGISDVIYNNFLL	1210
Db	1195	TRSHSGVQFMGPSSYVGCTGSSAMFSDAFNNSLIIMDYRPSLPRQL--KLITSPLTT	1251
QY	1211	AGEVVGXHRPS-----VIRKGNPDQNIQTLQAPPS	1241
Db	1252	RGPLINATGATSPASPTHRSPFNHLPRHPIPLSYDBDEFVYNTPLPDQGPETRLDT--HFLRP	1309
QY	1242	QOIAHYHQALAEELGHRAPAAVALPPCPREDEGLLYMQELTVSDSVLVFEL-----T	1294
Db	1310	SRPLHFHDLPPAI--PRPVSTSYVD--POAKASPYV-PEEF--FDSLAAFLPAHDPSTR	1353
QY	1295	DIVACRMAAPS-----QKAVLSTLVGRYGRRTKLYEAAH-----SDVRESIARFI	1340
Db	1364	EILKKDOSSNQFMWRDPRSLSC-----QPSLISAKIARPHDPTLLPASINRKL-RFR	1416
QY	1341	PTIGPVQATT-----C-----ELYELVAMVYKQGDGSAVLE	1372
Db	1417	PSDSPHQITADDDVVLGQLFHSICRAYSRQPNSTVPFNPFLPAECISILEYNAQLSSKTOS	1476
QY	1373	LDCICNRVYS-----RITPFQKXCKKFTTGETIAGKTKGQGISAMSKTFCALPBP--W	1422
Db	1477	TIVNASSRSPDWMRHTTVKITEFAAOKRVNDGSITFGSWKACQTLALMHDIYVILVIGRPVKY	1536
QY	1423	FRALEKEIILALPPNIF-----YGDAYEESVYAAAVSAGASCNVFENDESEFSDTO	1473
Db	1537	QRIDN---ADRPNNIYSHOGKTPNDRMQCEHL-----THSTPKIANDYTFATFDSQ	1586
QY	1474	NNFSGLGECVAMEECGPMQLIRLY-HLYVSAMIILOAPRESLKGFWKHSGEPGTLMMNT	1532
Db	1587	HGESVIVLEALKMKRLNIPSHLIQHLHVLTKNVSTQGPGLTCM-----RLTGGPGTYDDNT	1641
QY	1533	VWMAAIIIAHGYEPDRDFVNAFVKDDSDSVLCSQV--RQSNMAAALLAGGKIKLVXY----	1586
Db	1642	DYNLAIVYSQIDVQSGCPIW--SGDDSLI--DHPLPTRHDWPSVLKRLHLRKLELTJSHP	1697
QY	1587	-----RPIGLAVGVVYAPGLGTLPD-----VVRF--AGRLEKNNMGPGER	1625
Db	1698	LFCGYVYGPAGCIRNPLATLCKMLIAVDDALDRLSLVTFETTGHLGLGESHMLHPET	1757
QY	1626	AEOLRLAVCDF	1636

[illegible]

Db 454 -----SAYPASPPIQ 464
 QY 507 DVSQYAVHGHOLEALYRALNVPDIAARA-----SRLATYELVASPDR---551
 Db 465 KLP-----HALOKAALLILKIPSLTATPFERSEQSMJPNALSWLTKRFLPMQASL 519
 QY 552 -----LECRVYLGKTRRTYVDGAHLEANGPEQYVLSFQASROSQMGASHLYELTPAG 607
 Db 520 VILALSESSILHLKFSPPLOAOH-----DYYHRLHFGSYSLQWERP--564
 QY 608 LOYRISNGI---DCTATPPGAPSAAGEVAFCALYRYNRFOR-----HSLGLWM 660
 Db 565 LSLPRTATLPRTPTSTAPRDSASLP---PAFAST-----FYRPPRPAASPCA--613
 QY 661 LHDEGLLGP--PFSQGHMESANPCGEGTLYRTWSTSGFSSDSP-----P 708
 Db 614 -QPTTAAAPPTIEPTORAHQNSDLAESSTSEPPRPTIRSDMTPSAPVLFPEINSP 672
 QY 709 EAAPAPMAATPGI--PHSTPPVSDIWLPPPEEFOVD-----AAPVPAPD 753
 Db 673 RRPPLPAPTPLEPAHTPPPLS-----IPRODPTSDVPLMGSHLLHSLPAPPTPLPS 728
 QY 754 PAQLPPVVLTPPPPPV-----HKPS--IPPSNRRLTYT 788
 Db 729 SOLPLAPLINDPTAIGPVLPFEELHRRRYRPPENTATFLTLRLSLPSNHLPOPTLN-CLLSA 787
 QY 789 YPDGARVYAGSLFESDCDMLVNA--SNPGHRRPG-----GLCH-----AFYQRFPE 832
 Db 788 VSDQTVSEHLMESLOTLIPDSQNSNETNLTGLSTEHLLTALAHLYNQATVYSDRGPI 847
 QY 833 APTPTFFIKREGLAATTLIPRIIHAVAPRYREONPKRLEAAYRETCSKRGTAANPLIG 892
 Db 848 LEPSPOTIKR-----IDTHTTGPSPHSPG-KRLGSGP--SAKGHSPDLIR 893
 QY 893 -----SGIYQVSLSDAWERNHRPGDELY---LTPEAAMFWFANKPAQVLTITTED 942
 Db 894 AKSEKVSQNY-LPFS-----FAHHPTISIAKMLISMKNKGFQVSLDVSIGQRT 946
 QY 943 TATFANLALDATEVGRACAGCTISPG---IVHYOFTAGVSGSKSIQO-----992
 Db 947 GPTPKRIIQIDHYLDT-----NPGKTPVVFH---AGPAGGKTPYQQLKTKI 994
 QY 993 -GDVYVYVYTRRLKSMKR---RGFAA--TPPTAARTYTRRYVIDEAPSLPHLLI 1045
 Db 995 FKDFR-VSCPTELRLREKMTAMELHGSQSMRFMTWESSILKSSRIIVIDEIYIMPGYLD 1053
 QY 1046 LHMORASSYHL---LGD-----NQIPADIEHAGLIPATIRPELAPISWMX 1088
 Db 1054 LSLADPALLELYIILGDPLOGEYHSQSKSSNHRPS---ETIRLPLTY---DMTCMW-1105
 QY 1089 VTHRCPADVCELL-----RGAVPKIQT-----TSRVLSLSEFMNEPAIGOKLVXT 1132
 Db 1106 -SVRIPOCIARLEQIHFSNANMGVIGSVSTPHDQSPVLINSHASLSLTFN--SLGTYR---1158
 QY 1133 QAKAANPCALIVHEAQATFETTTIATADAGLIOSSRAHAVALTR-----1181
 Db 1159 -----SCYISSQGLTFCDPAIIVDNTKWLSS--ANGVALVATRSRSGVQPMGPS 1207
 QY 1182 -----HTEKCVILDA--PGLAREVGISDVIYNNFLAGEVGXNRP---1220
 Db 1208 STYVGCTNGSSAMFSAFNNSLIIMDKYFSLPQL---KLITSPLTTRPKLNGATPSAS 1264
 QY 1221 -----SVIPRGNDONLCTLOAFPPSCQISAVHOLAEEL 1254
 Db 1265 PTHRSBNFHLPHIPLSYRDVYVNSSTLPDQGPETRLT--HFLPRLPLHFLDLPPI 1322
 QY 1255 GHRPAPVAALDPCEPLEGLLYMDELTVSQVLFEL-----TDIVHGRMAAPS--1305
 Db 1323 --TTPPVSTSVDP--PQAKASPYV--PGEF--PDSIAFAFLPAMPSTRELHMDQSSNQRP 1376
 QY 1306 --QRAVAVSLTVGRYGRKRYLEAAH-----SDVRESIARFIPTIGVQAT--1350
 Db 1377 WEDRFSLISC-----QPSLSISAKHAPNHDPDLLPASINKRL--RRPSPDSPHQITADOV 1429

QY 1351 -----C-----ELYELVAMVERKODGSAVLELDCNRDVS-----1381
 Db 1430 VLGLQLEFHSICRAVSRQPNSTVPFNPFLFAECISLNEYAQLSSKSTQSTIVANASRSDPW 1489
 QY 1382 -----RIFFQKXCKKFTTGETIANGKYGQGISAMSKFQALFGR--WFRAIEKEILLALP 1435
 Db 1490 RHTTVKIFAKAQAQHKVNDGSIJFGSKACQTLAMHDVYIILVGVKKRYQRIEDN--ADRP 1546
 QY 1436 PNIF-----YGDAYEESVFAAASVAGSCMVFENDESPDSTQNNFSLGCVME 1486
 Db 1547 PNIIYHCGKTPNDRMCOEHL-----THSTPKIANDYTAPDQSGHGSVYLEALKK 1599
 QY 1487 ECGMPQMLRLY--HLVRSAMILQAPKESLKGFKKHSGEPTLLMNTVMNALIAHCYEF 1545
 Db 1600 RLNIPLSHLIQHLVHAKTNVSTQCPILCM-----RLTGERGYDMDNVMIAVYSQYDV 1654
 QY 1546 RDRVVAAFKGDSDSVLCSDY--KQSRMAALLINGCGLAKVDY-----1586
 Db 1655 GSCPIWV-SGDSLI---DHPILPTRHDWPSVLKRLHREFLELTSHPLFCGYVYGAGCI 1710
 QY 1587 -RPIGLYAGVVAAPGLGTLPD-----VVRP--AGRLSEKNMGGPPRABQLRLAYCDF 1636
 Db 1711 RNPLALFCKLMIADVDDALDRLRLSTLFTPTTGHLLGESLMLHLLPETHVQYSACPDF 1768

 RESULT 10
 RPO.ORSVC
 ID RPO.ORSVC STANDARD: PRT: 1612 AA.
 AC P89659; 039640;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [COMTAINS:
 DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
 OS Odontoglossum ringspot virus (isolate Korean Cy) (ORSV-Cy).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 NX NCBI_TaxID=138661;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96381046; PubMed=8789059;
 RA Ikegami M., Isomura Y., Matsumoto Y., Chatani M., Inoue N.;
 RT "The complete nucleotide sequence of odontoglossum ringspot virus (Cy-
 RT 1 strain) genomic RNA.";
 RL Microbiol. Immunol. 39:995-1001(1995).
 CC - FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 CC RNA REPLICATION.
 CC - FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
 CC CAPING AND AN RNA HELICASE.
 CC - MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
 CC BETWEEN CODONS FOR LEU-1112 AND GLN-1114.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D13941; BAA21828.1;
 DR EMBL: S83257; AAB94948.1;
 DR InterPro: IPR001788; RNA_dep_RNApol2.
 DR InterPro: IPR002588; Y_methyltransf.
 DR InterPro: IPR000606; Viral_helicel.
 DR Pfam: PF00978; RNA_dep_RNApol2; 1.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Ymethyltransf; 1.
 KM Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
 FT CHAIN 1 1612
 FT NP_BIND 826 833 METHYLTRANSFERASE/RNA HELICASE.
 FT ATP (POTENTIAL).
 FT SEQUENCE 1612 AA; 183252 MW; 3F803A2B9611EDDF CRC64;

Query Match 3.8%; Score 339.5; DB 1; Length 1612;
Best Local Similarity 19.2%; Pred. No. 8.7e-14;
Matches 351; Conservative 232; Mismatches 659; Indels 585; Gaps 85;

```
QY 94 LHHCF-----LRPGRDQRYYSAPTRGPANCRSALRGPR-----PADRTYC 137
DB 123 VHCMPMLDIRDYARHINODTVSTY--LARLEKSK-RGLPRVQOQSAPNKMYMDPDVACC 179
QY 138 FDFGSRCAFAET-----GVALYSLHDLMPAD--VAEAMRHGXRTLAALHLPPEVLPP 191
DB 180 DKRPQDQSVLDLPGKTYAVALHSTYDI-PADERGALLKRDVHICAAHFSNLL-- 236
QY 192 GTHTTSTYLL-----IHGDRAVVTYEGDTSAGVNHVDSILRAMTIT--KIVG 238
DB 237 ---ETTSAPDEIGATPFYKSGDRLSEFFQNESTLNTYSHSKNVIKYCKTFFPASNRFVY 293
QY 239 DHPFLYERVAIGHFVLLLT-----AAPSPMPYRP----- 271
DB 294 HKRPMCTRYVTFWCKTKVVTYFLFKGVYTRGEDSEOFYAMDEAMEYKKTTLAMLKCEPT 353
QY 272 -----YPRSTEVYRSIF-GPGGSPSLFPSSACSTKSTF-HAAPHVIMDRMLFG 318
DB 354 IFDRBAAVNFMFPKVKMDVIVPLFDGSVTSGKKMRSEVMNKNDEYTVLNI----- 405
QY 319 ATLDQAFCSSRLMTYLRGISYKYVYALVANGMNASDALTAIXITAAVLTICHQRYLR 378
DB 406 RTYQDKALTYKNVLSFESVIRSRVYIINGVSAREMDVKSVLQA-LSMTEFL-----LQ 457
QY 379 TQMSKGRMLGVEHNOKF---ITRLYSWLFESGSDYITGRQQLFAQCRRLWSAGFHL 435
DB 458 TK-LAEKKDOVYLVAKFEKEDVTNLEKWOISDAVGDLPSPK-----ETLISGPF-- 507
QY 436 DPRVLVDESVPQCRCTFLKVKAGKFCFCFMRLGQECTCELEPAEGLVGDHNDNEAYEG 495
DB 508 ---VKVAEQLQI-----KTPDEYITF-----ADKLIME-----YKA 536
QY 496 SEVDPAEPALHDVSGITYAVHGHQLEALYRALNPDIAARSRITATVELVASDRLECR 555
DB 537 TE---ELQHLDISKPL---ERAEKYVNL-----SELSTVUKE-CDEFDIYQFK 577
QY 556 TVLGNKTFRTTV-----DGAHLEANGPEQVLSFDASRQSMGAGSHLLEYLTPA 606
DB 578 NICEEKDIDDVYAKVIALMKNELTLPKNPPEALS-DA-----LSLPLK 623
QY 607 GLOVRISSNGLDCTATRP-----PGAPSAAPGEVAFCSSALRYNRFTQHSLSLGLW 660
DB 624 DLMRFDLTKLSTCAPRPPSVYKTLDSGLLPKQSYGDERQF-----ESQSVASVD 672
QY 661 LHEDGLGITPPSPRGHIMSANPFCGEGILYTRTMTSGFSSDFSPREAAAPAMATPG 720
DB 673 FHLKSV-----BSVKMKSMSAAVYTGRLVQOKNMY--DYLSASISAT-- 714
QY 721 LPHSTPVSIDWLPPESEEFQVDAVPAPRDPAGLPGVVLPRPPRPVHKPSIPPS 780
DB 715 VSNLCVKLKVYGDAPESAE----- 734
QY 781 RNRRLTYTPDGAQVYAGSLFESDCMLVNASNGHNRPGGGLCHAFYQRPPEALYPTFEI 840
DB 735 --KSGVYDVVKG-----KWLTKPK-----GKCHAM----- 757
QY 841 MRGLAAYTLTPRPIIAVAVPDRVEONPKRLAAYRETSRKRTAAYVPLIGSITYOVPV 900
DB 758 --GVALNNGEKVY-----LLEMA-----DGFPIIGD----- 783
QY 901 SLSEFAMERNHRHPRDELYLTE--PAAMWFANKPAQVPLITTEDTARTANLALIEDATE 958
DB 784 -----WRRVAVSSDSLITSDMKLQTLUSCLKDGEVY----- 815
QY 959 VGRACAGCTISPGIVHYQTAGVPGSGKRSIOG--DVDVVVVPTRELNSWRRRG-- 1013
DB 816 -----PSDAKVTLVGDVPGCGKTKEILETVNFDEDLVLIVPKKACMKIIRKANK 864
```

```
QY 1014 -----FAFTPHTAARVTIGRRVYIDEAPSLPHLLH-----MORA 1051
DB 865 SGHVRATKDNVRYTDSFLMLHKPTY--NKLFEDEG-----LMLRTGCYNLIALSHC 915
QY 1052 SSVHLGDPNOIPAIID-----FEHAGVPAIRPLAFTSMWXVTHRCPAD 1096
DB 916 REAMVEDTQIPIFINVANFPYKPHATLVYDHR-----EVR-----RLSLRCPAD 962
QY 1097 VCELIRGAYP-KIQGTTSVLRSL-----FVN---EPAIGKLVYQOAKA----- 1137
DB 963 VTHRMNSKYDGKVLCTNDVIRSVDAEVYRGKGVNPKSPKPKGIITFTOSDKKELKERG 1022
QY 1138 -----ANGAI-TYHEAGATFTETII-ATPADAGLIQSSAAHIVALTFRTER-----CV 1187
DB 1023 YEEVSTGEINFTYHEIGEFEDVSVRLPPTPLELISKSSPHVALVLTFRTKSEKYYSV 1082
QY 1188 ILDAPGLLEVG-----ISDVIVNNEFLAGEVGHRPSVIPRGNDONLGTQAFPPSCQ 1243
DB 1083 VLDP--LVKCSDSLKSYSDFTLDMKYKADAGILXQLOVGSIFKG---ENL-----FVP-CP 1131
QY 1244 ISAYHOLAEIGHRPAPVAAVLPCCPEL--EQLLYM---PQLTVSDSVLVFELTDIVH 1298
DB 1132 KSGY-----ISDMQFYFDTLPGNSTILNEYDAVTMLRENNINVRXDCITDFSKS---- 1181
QY 1299 CRMAAPSORKAVLSTLVGRYGRRTKLYEAHNSDVRESLARIPT--IGPVQATTCLEYEL 1356
DB 1182 --VSVPQOQEEFFPVT-----RTAERPRSGGLENIIVAMIKRNPSPDLTGILDEDT 1234
QY 1357 VEAMVEKGODSAVLELDICNDVSRITTFQKXG---NKFTGETIAHGKVG----- 1405
DB 1235 AELVYKMFWDAYITIDELSGGVNTPTSDAFIRMAKQEKSTIGLADPDFDLPAIDQYK 1294
QY 1406 -----OGISAMSKTCALFGPWRFAIEKEITALLPPN--IFYG 1441
DB 1295 HMIKAPQKQKLDLPODEYALQTIYVHSKQINAFGLFSELTROLIERIDSSKFLFYT 1354
QY 1442 DAYEESV--FAAASGAGSCWVFENDSEPSQTQNNSTSLGECVYMEGCPQWLLIRYH 1499
DB 1355 RKTPEQIEEFPSDSDTSVPMKVELDLISKYKSONERCAVEYLIWEVLGNGFL----- 1409
QY 1500 LVRSAMILQAPKESLKG-----WKHSGPGLTLMNTVNMMAIIAHCEYFRDFRV 1550
DB 1410 --EEVMQGHKRTSLKQYTAGIKTCIMQYRKSGDVTFTEIGNTVIAACIASIMPMKYK 1467
QY 1551 AAFKGDSDVYLC--DYRGRNAALIAAGGLKLYKDYRP-IGLIYAG-VYVAGGLGLTP 1605
DB 1468 AAFCGDDSMLYIPKGLDLPDIOSGANLMMNFEAKL--YRRRYGVCGRYIIHHDRAIV 1524
QY 1606 --DYVRFAGRISEKNMGVGPERRAOLRLAVCDPLRGLTNQV--CYDVVASRYVGVSPGL 1661
DB 1525 YDYPVKLISKGCKHIT-KSLDHLIEFRISLDCDVASNLNCAVYQOLDATAIEVH----- 1577
QY 1662 VHNIGMLQTIADGKAHETETIKPVL 1668
DB 1578 -----KTAVNGSPAFCSIVKTYLSD 1596

RESULT 11
RRPO_TOML
ID RRPO_TOML STANDARD; ' PRT; 1616 AA.
AC P03587; 041352;
DT 21-JUL-1986 (Rel. 01, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tomato mosaic virus (strain L) (TOMV) (TMV strain tomato).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=12252;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85157522; PubMed=6549393;
```

RA Ohno T., Aoyagi M., Yamanashi Y., Saito H., Ikawa S., Meshi T.,
 RA Okada Y.;
 "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
 RT genome and comparison with the common strain genome.";
 RT J. Biochem. 96:1915-1923(1984).
 CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 CC RNA REPLICATION.
 CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
 CC CAPPING AND AN RNA HELICASE.
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
 CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X02144; CAA26085.1; -
 DR EMBL: X02144; CAA26082.1; -
 DR PIR: A04195; WMTM8T.
 DR InterPro: IPR001788; RNA_dep_RNApol2.
 DR InterPro: IPR002588; V_methyltransf.
 DR InterPro: IPR000606; Viral_helicase.
 DR Pfam: PF00978; RNA_dep_RNApol2; 1.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Vmethyltransf; 1.
 DR Transferrase: RNA-directed RNA polymerase; Helicase; ATP-binding.
 KW CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
 FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
 FT NP_BIND 833 840 ATP (POTENTIAL).
 SQ SEQUENCE 1616 AA; 183564 MW; A8EC929B5CF7CAFC C64;

Query Match 3.6%; Score 322.5; DB 1; Length 1616;
 Best Local Similarity 19.2%; Pred. No. 1.1e-12;
 Matches 349; Conservative 205; Mismatches 660; Indels 601; Gaps 79;

QY 91 PNLHLCFLRPVGR--DVQWYNSAPTRG--PANCRRSAL-RGIPRADRTYCCDFGSRCA 145
 DB 128 PNLVDNDIMHNGOKOSIELYLSRLERGNKNVNFQKAEADRYAEPMENEVVCCHDTFOTCR 187
 QY 146 FAEE--TG---VALXSLHDLMPAD--VAEAMARHGXTRLAALHLPREVLLPG--TYHT 196
 DB 188 HSECECTGRYATLALHSIYDI-PADERGALLKKNVNCVCAAFHFSNLLDESHVWDE 246
 QY 197 TSYLLHDCRAVYVEGDTSGAGYNHDVSLIRAMIRTTKIVGDHPLVIERVRAIGCHEVL 256
 DB 247 INACFORDCGRILTFEFASESTLNVSHSYNLLKYVCKT----- 284
 QY 257 LITAAEPSPMPYVPRSTREYVVRSLFPGGSGPSLFPSACSTKSTTHAVPVHIMRLML 316
 DB 285 -----VF-P-ASNEEYVWKE-----FLVTRVMTW----- 306
 QY 317 FGATLDDOAF-C-CSRIMTYV--RGISKYVAGALVANEGNASEDAATATATAYLTICH 373
 DB 307 -----FCKRSRIDTFLLYKGVAHK-----GVDSEQFKAMEDMANYKTT---LAMCN 350
 QY 374 -ORYLTQALISG-----MRLGLEVAOKFITRLYSWLFESGR-----DIYPRQ 418
 DB 351 SERILLEDSSSVNYWEPKMDM-----VIVLEFDSILETSKTRKREVLVSKDFV----- 399
 QY 419 LQFYACR--RMLSAGFHLDPRLVLDSEVPCRCRTFLKKNVAKFCFCFMWLOQECGFL 476
 DB 400 ---TYVLNHIRYQAKALYTSNVLSFEVSI--RSRYIINGVTAH----- 428
 QY 477 EPAEGLVGDHGDNEAYEGSEVDPAPAPHLVDVSGTYAVHGHQLEALYRALNVQDIAARA 536
 DB 439 -----SEMDVDSKLSLOSLSMTFELH----- 458
 QY 537 SRLTATVELVAPDRLECRIVLGNKTRPTTVVDGAH----- 572

DB 459 -----TKLAVKDDLISKALGPRTVSOHWMEISLAFGNAPSIKERLINRKLKITE 513
 QY 573 --LEANGPEQYVLSFD--ASRQSGAGSHSL-----TYELTPA--GLQVRISNGLD 618
 DB 514 NALFIRVPLDYTFHDLVSEYKMSVDMPVLDIRKKEEFDEKMNALSELVSKNSKFD 573
 QY 619 CTATFPFGAPSAAPGEVAAFCGALYRNYNFTORHSLTGLMLHPEGLGIIPFPSPGHI 678
 DB 574 VDV-----FSQMCQSL----- 584
 QY 679 WESANPCGEGCTLYTRKWSISGFSQSPSPPEAAPAMAATPGLPHSTPPVSD--IWLPP 736
 DB 585 --EVDPTTAQKVIYAVVNSNSGTLTFPEOPTANVALAL-----QDEKASDGAALVYTSR 637
 QY 737 PSEFQVDAPVPAPADPAGLPQVYVLTTPPPPVVHKPISPPSRNRLLYTDPDGAKY 796
 DB 638 DVEBPSIKSGMARCELOLAGSDV-----PES-----STRSEIE 674
 QY 797 AGSLFESDCDMLVNASNPGHRPGGGLCHAFYORPEAFYPTFEITMRGLAAYTLTPRPDI 856
 DB 675 SLEQFH-----MATFASLLIKQ--MCSIYTTGGLKQGMKNFI--DSLVAASLSAAVSNL 724
 QY 857 HAVAPDYRVQNPRLKLEAAVRETCRRGTAYPLLSGIIQVPVLSFDAM--ERNHRPG 914
 DB 725 VKILKTPAIDLETRQKFGVLDVASKR-----VLVPSAKNHAMGVETHARK 772
 QY 915 DELTYNE-----PAANWEANKPAPVLTITEDTARTANLLEIDAATVEGRACGCT 967
 DB 773 YHVALLEHDEFGIITCDNMNRVAVSSSVYSDMAKIKTRLRLK-DGEHVSSA----- 826
 QY 968 ISPGIYHYOPTAGVPSGSKRSISIQG--DVVVVVVYPTRELRSNRRR--GEAPTPT 1020
 DB 827 -----KVLVLDVPGGCKIKELISRYNFEEDLILVGRQAAEKIRRAVNSGIIVATKD 880
 QY 1021 TAARV-----TIG-----RRVVIDEAPSLPHLLH-----MORASSVHLGD 1059
 DB 881 NRTVVSFLNNYKSGARQCFKRLFDG-----LMLHTGCVNFLVEMSLCDIAYVGD 933
 QY 1060 PNOIPALDEFHAGLVPA--TRPELAPTSMMXVTHRCPADVCCLRGAYP-KIQTSVLR 1116
 DB 934 TQOIPYINRYTGPPYRHNFKLEDEYETRTTLKPCADVTHTLQNRBEHVMTSSEKK 993
 QY 1117 SLFW-----NEPAIGOKLVYTOAKAA-----NPGATVHEAGATFTETII 1159
 DB 994 SVQGEVMSGAASINPYSKRLKGIILFTQSDKEALLSGVADVHTVEVGEFYADVSLV 1053
 QY 1160 -ATAARGLIQSRAHAVALTRHTEK-----CYILDA-----P 1192
 DB 1054 RLTPPTVYSILARDSPHVLVLSLHRTKSLKYTVVMPPLVSIINDLERVSSYLDMYKVA 1113
 QY 1193 GLIREVGISDIYVH-NPEFLATGGEVGMHRPVIIRGNDDOVLGTLQAPPPSCQISAVHQLA 1251
 DB 1114 GTOXOLOVDSVEKFNFLFVAPKGTG-----DISDQFYDKC----- 1150
 QY 1252 EEELGHRPAPVAAVLPCCPELEOGLLYMPELTVSDS--VYVFEITDI-----VHC 1299
 DB 1151 -----LPGNSTILNNNDATYMKLTJDISLVWKCILDMSK 1184
 QY 1300 RMAAPSORKAVLSTLVGRGRRTKLYEAHSDVRESIARTPT--IGPVQATTCELYELV 1357
 DB 1185 SVAAPKPDVKTPLTPMV-----RTAAEMPRQOTGLENLVAAIKRNFSPELSGVVDIENTA 1239
 QY 1358 EAMVVEKGODSAVLELDLCNRDVSRTFFQXCNKF--TTGETIAHGKVG----- 1405
 DB 1240 SLVYDKFFDSYLLKEKKRPKNFS--LFSRESLNKMAKOEQYVITIGOLADFDVVDLPAYD 1297
 QY 1406 -----OGISAMSKTFCALFCGPMWRAIEKELLALL----- 1434
 DB 1298 QYRHMIAKQPKQKGLDLSIQREYALQTIYVHSKINAIIFPRLPSRLRLDLSIDSSRFL 1357
 QY 1435 -----PPNLT--FYGDAYEESVFAAAYSGAGSCMVENPDFSSEDSYQNNNSLGLCECYME 1486

CC	-1-	MISCELLANEOUS:	READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC			
CC		BREATHING CODONS FOR GLN-1116 AND GLN-1118.	
CC			
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC		the European Bioinformatics Institute. There are no restrictions on its	
CC		use by non-profit institutions as long as its content is in no way	
CC		modified and this statement is not removed. Usage by and for commercial	
CC		entities requires a license agreement (see http://www.isb-sdb.ch/announce/	
CC		or send an email to license@sdb-sdb.ch).	
CC			
DR	EMBL:	AJ243571; CAB62911.1;	
DR	EMBL:	AJ243571; CAB62912.1;	
DR	InterPro:	IPR001788; RNA_dep_RNApol2.	
DR	InterPro:	IPR002588; V_methyltransf.	
DR	InterPro:	IPR000606; Viral_helicaseI.	
DR	Pfam:	PF00978; RNA_dep_RNApol2.	
DR	Pfam:	PF01443; Viral_helicaseI.	
DR	Pfam:	PF01660; Vmethyltransf.1.	
KM	Transitase:	RNA-directed RNA polymerase; Helicase; ATP-binding.	
FT	CHAIN	1 1616	
FT	NP_BIND	1 1116	
FT	CHRAIN	1 1116	
FT	MR_BIND	833 840	
SQ	SEQUENCE	1616 AA; 183606 MW; 4BF913CE791DE33 CRC64;	
	Query Match	3.6%; Score 320.5; DB 1; Length 1616;	
	Best Local Similarity	19.3%; Pred. No. 1.5e+12;	
	Matches 350;	Conservative 223; Mismatches 646; Indels 591; Gaps 83;	
OY	91	PNVLRHCFRLPVGR-DVQRWYSAPTRG--PAANCRRSAL-RGLPRADRTYCFDGSFRC	145
Dd	128	PNLVVRDIMEHGCKDSIELYLSRLERGNKHVPNQKEAFBRYAEMPNEVCCHDFQICR	187
OY	146	FAAE-TG---VALYSIDLMPAD-VAEAMARHCXTRLUALHLPREVLPRG--TYHT	196
Dd	188	HSQECYGRVAYIALHSTYDI-PADEFCAALLRKRVHVCYAFHFSENLLSDSHVNIDE	246
OY	197	TSYLLIHGDRAVVYTEGGDSAGYNHDVSIILAWITRTKIYGDDHPIETERRAIGCHEVL	256
Dd	247	INACFORGDRLTFSPASESTLNKSHSYSLIKLYCKT-----	284
OY	257	LTLAAPESPMPYPYPPYSTEVYVASITFGPGSGPSLFPSACSTKCTFHAVPAPHIDRLML	316
Dd	285	---YTP-ASNRREVYKME-----PLTVRNTW-----	306
OY	317	FGATLDIOAFC-CSRIMLYL--RGISTKYVGALVANEG-MNASDALTAIXTTAAVLTIC	372
Dd	307	-----FCKFRIDIFELLYKGAHNCG-----VDNQFPKAMDANHMYKT-----LAMC	349
OY	373	H-QRYLTRQAISKG-----MRLGEVHAOKITRLIYSLWFLEKSGR-----DYDER	417
Dd	350	NSERIILEDSSSVNYWPFKRDM-----YIVPLFDISLETSKTKREVLVSXDFV---	399
OY	418	QLQFYAACOR--RWLSAGPHLDPRYLDPEDSPBCRCRTLFLKYAGKFCFPMRLAGECTCF	475
Dd	400	----YYLVNHRTYQAARKALTYSNVLSPEEST--ISRVIINGVTAR-----	438
OY	476	LEPAEGLVGDHGHNDNEAYESGEVDPAPBAHLDVSGTYAVHGHOLEALYRALNVPODIAAR	535
Dd	439	-----SEMVDKSLDLSMTPLFH-----	458
OY	536	ASRLTAIVELVASDRLECRVTGLNKPTRTTYVDGAIHLEANGPREDYVLSFDSRSMGAG	595
Dd	459	-----TKLAVLKDDLISKPALGPKVYSQHVMWEISLATFN-----AFPSIKIRLL--	503
OY	596	SHSLTYELTPAGLOVRILSNGLDCTATPCPGCAPSAAGVEAAFCALYRNRFQRHSL	655
Dd	504	INKRLKITTEALEIRVP---DLXYTP-----HDRLYSEYKM	537
OY	656	TGGIWL-----HPREGLGIPPPSPGHIVESANPFCGGLTYLRTWTSGFSDFSPPEA	710
Dd	538	SVDPVVIDIRKKMETEEEMYNALSELSTVLTKSDSFEDV-----VFSONCOSLEVDPMIA	591


```

Db 934 PYAHAEKLVVDEKEDR-----RVTLRCPADVYFFLNOKYDGSVLTSSVSERSVSAEV 987
1123 -----PAIGKLVVYTOAK-----AANGALTVHEAGAFTEETITII-ATADAR 1165
Db 988 RKGALNPITLPLEGKILTTTQADKFEELDKGYDVTVEHVGSETETKTAIVLAIPL 1047
1166 GLIOSSRAHAIVALTTRTEKC---VILD-APGLIREVG-ISDVIY-----1205
1048 EIISSRAPHVVALTRHTTCKKYTVVLDPMVNVISLGLNSFLLEMYVESGTQXQ 1107
1206 -----NNFFLAGGEVGHXR-----PSYIPKGNP-----DONLGLTQAFPPSCQ 1243
1108 IDTFVGTNLTFVPPKSGWRDMQFYDTLLP-GNSTIIFEFAYTVNMLDISLVNDCR 1166
1244 I--SAYHOLAEELGHRPA-----PVAALVPCPELEFOGLIYM-POELTFVSDSLVELT 1294
1167 IDSRSKVQVPE---RVEFMKPKLRTAEMPRTAGLLENVAMIKRMNAPDLTGTDIE 1223
1295 DIYHCRMAAPSOKRAVILTVGRYGRRTKLYEAHSD---VRESIARFT----PTGP 1345
1224 D-----TASLVVEKFWADYVYKEFSGETDMAMTRSEFSRWLSKQESSIVGQ 1269
1346 VQ-----ATGCELVYELVAMVEKGGDSAVLELDLCNDRVGRITFPCKXCKFTTGER 1398
1270 LAFNPNVDPADVEYKHKMISQPKQ-----KLDLSIDQ-----EYPALQT 1309
1399 IAHKVGOGISAMSKFCALFCPFWFAIEKEIILALPPN--IFYGDAVEESV--FAAAS 1454
1310 IYV-----HSKRIALFQMFSELTMLLETIDTSKFLFYTRKTPQIEEFSDLD 1360
1455 GAGSCAWFEDESEFDSTONNESLGLCEVYMEOGMOWILRLYLHVSAMIIQAPKESL 1514
1361 SSQAMEILLEDISKYKQSEHFCAYEYKIMERKIGIDDLAEV-----WROGHKRTTL 1413
1515 KGF-----WKHSNGEGLTLMNTVMMAIIAHCEYERDPRVAFKGDSDSVLCSX 1565
1414 KDTTACIKTCWYKQSGSDTFTTIGNTIIIAQLSSKIMPMKVIKAFCDSDSLIYIP-- 1471
1566 ROSRNAALTAGCGL-----KLKVDYRPIGLYAG-VVYAPGLTLP--DVYRFAGRISEKN 1618
1472 -KGLDLPDIOGAGNLTMWNEAKLFRRKKYGFYGRYVYIHHRGALIVYDPLKLSKLGCKH 1530
1619 WCGPERABOURLAVCDFELRGLTNVAOCV--DVYSRVYGVSPGLVHNLGLQTAD 1674
1531 I-RDEVHLEBELRSLCDVSNLNNCAVFSQLEDAVAEVEHTAVGAFVYCSIIKYLSD 1587

```

RESULT 15

```

RPO_TM0B
ID RPO_TM0B STANDARD; PRT; 1616 AA.
AC P90211; 083484;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
  METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tobamovirus OD.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=31749;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93189450; PubMed=8376970;
RA Ikeda R., Watanabe E., Watanabe Y., Okada Y.;
RT "Nucleotide sequence of tobamovirus Ob which can spread systemically
  in N gene tobacco.";
RL J. Gen. Virol. 74:1939-1944(1993).
CC - FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
  RNA REPLICATION.
CC - FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
  CAPPING AND AN RNA HELICASE.
CC - MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
  BETWEEN CODONS FOR ALA-1115 AND GLN-1117.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D13438; BAA02700.1; -
DR EMBL: D13438; BAA02701.1; -
DR InterPro: IPR001788; RNA_dep_RNApol12.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR000606; Viral_helicase.
DR Pfam: PF00978; RNA_dep_RNApol12; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Ymethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616
FT NP_BIND 832 839
FT NP_BIND 1 1115
FT NP_BIND 1616
SQ SEQUENCE 1616 AA; 182997 MW; 146BD679AA46575 CRC64;

Query Match 3.5%; Score 319.5; DB 1; Length 1616;
Best Local Similarity 18.1%; Pred. No. 1.7e-12;
Matches 329; Conservative 218; Mismatches 580; Indels 693; Gaps 74;

OY 132 ADRTYCEPDGSRSC---AFAETGVALYSLHDLM--PAD-VAEAMARHGRLTYAALHLP 185
Db 174 SDEVVCNNTFOCCESNRYSSGGRYVALSLSLDIPADEGALMLRNHTLTYAAHFPAE 233
OY 186 EVALLPGTYH--TTSYLLIHDGRAVTVYEGDSAGYNHDVSLIRAMIRTT-----KIV 237
Db 234 ELLELVSTVELPTIGGIFSDGDKINFCSENESTLWNSHSLKLYVCTTPPASNRVY 293
OY 238 GDHPLVIERVRAICGHVLLITAAPESPMPYVPYPRSTEVYRSIFPGCSPLFPSPAC 297
Db 294 YKKEFLITRVNWFCKFTKIDT-----YLLY-----RGVYHRCQDOQEYSAM 336
OY 298 ST---KSTFNAV-----PYHINDRLM-----LFGATIDQ----- 324
Db 337 EDAMHYKKTILAMLSERIVLEHSSVNVWPPKKDMVIVPLFVLSLETOQRKREYISK 396
OY 325 -----AFCCSRMTYLRGISYVVTGALVANGMA----- 355
Db 397 DFVYTVLNHRTYQAKALYNNVLSFEVSTRSVIINGYARSEMDVDALLOSMAMTF 456
OY 356 -----SEDALTAXITAAVYLT-----CHORY-----LRTQATSKGMRL 389
Db 457 LITKLSMLKDELIVSKFTLSAKSVHEHVMDEIKRGCNNPSPLESLLRKKLISGSAEL 516
OY 390 GVHNAKFTITRLYSMLFEKSGRYITGRQL-QYACCRKWLSGPHLDPRVLYFDESV 447
Db 517 EIEVPMYVTFHDFVFAEYKASVEMPTIDISKLSAEYSALSLS-----VLENSKDF 572
OY 448 -----CRCRFLKLVAGKFCFCFMRWLGQECTC---FLEPEGLVGHGHNEYE- 494
Db 573 DLEKFSRMAINCINVDIAKIVAV--LSNEGTVLLPKEPEEGMAAMKSGEDEV 630
OY 495 --GSEVDPAPAPHALDVSGTYAVVHGQLEALYRLANVPDIAARASRLTAVELVASPRL 552
Db 631 TLGSGQDNLT-----DLTSSMYTSSSLPLCGIASEL 661
OY 553 ECRVYLGKNTFTFTVVDGAHLNANGPEQYVLSFDSRSGMGAGSHSLYTELPAGIQVRI 612
Db 662 SCDTFVRNE-----EINSLIEYHML--AAESYISNMKASIVYS--GPLYOYQ 704
OY 613 SSGNLDCTATFPFGGAPSAPEV-----AAPSALYRNRNFRORSISLGLMLHE 664
Db 705 MONYDLSIA-----ASLSTVSNLKLKLVKDVSSVGFQDSLKGVDFVRK---MWL--- 752
OY 665 GLIGTFPPSPGHIMESANPFCGEGTLVYRTWSTSGFSSDFSPPEAAMATPGLPHS 724

```



```

Db 753 -----IKPFLKNSWGVQKFDGKCFLLALSY-----HN 781
QY 725 TRPVSDI-WLPPSEEFQYDAAPVPADDPAGLPGHVLTTPPPPPVHKPSIPPSRNR 783
Db 782 ELPICDADWSKVAVSNESWY-----801
QY 784 RLITYPDGAKYVAGSLFESDCDWLVNASNPGHRPGGLCHAFYQRRPEAFPTFETIMRE 843
Db 802 -----TSDMAKL-----808
QY 844 GLAAVTLTPRLIHAAVPDYVEQNPKELEAAYRETCSRGRGTAAYPLGSGIYOVPSLS 903
Db 809 -----KVLKSGEMPISTVS 823
QY 904 PDAMERNHRPDELVLEPAANWFANKPAQVLTTEDTARTANLALETDAATEVGRAC 963
Db 824 -----SAKVTL-VD-----831
QY 964 AGCTISPGIVHYQTAGVPSGSKRSIQG---DUDVYVYPTRELNRSMRRGRFAATPH 1020
Db 832 -----GVPGCGTKTELRRVNESEDLVLPGRKEAAMIRKR--ANQSGN 873
QY 1021 TAAR-----VTIG-----RRVIDEAPSLPHLL--LLHMORASSVHLLGDP 1060
Db 874 IVANNNDNKTVDSEFLNMLGKPGVCQFRRLFPVDEGLMLHPCYVFLVSLCNEAFVFGDT 933
QY 1061 NOIPAIPE-----EHAGLVPAIRPELAPTSMWXYTHRCPADVCELIRGAYP-KIQTTTS 1112
Db 934 QQIPYINRYQNFPPQHFSLIYDETEKRRT-----TLRCPVDYTHFLNQCYDGAVTTS 988
QY 1113 RVLRSLFW-----NEPALGOKLYXTQAK-----AANPGALTYHEAGATFTE 1155
Db 989 KTRGSVGLLEVVGAAVMNPYTKPLKGIYFTQSDKLMLSRGYQDVNTVHEIOGETYEE 1048
QY 1156 TTII-ATAARGLIQSSRAHAIALTRTEKC-----VILD-APGLREVISDVIYNF 1208
Db 1049 VSLVRLTPPTIHIISRESFHVIGLIRHT-RCKRYTYVLDPLVKLYRDL-----ECYSNF 1103
QY 1209 FL-----AGGEVGHXHRPSVIPRGNDNLGTLQAFPPSQOISAYHQLAEE 1253
Db 1104 LLDVYVWDSAXQLOVSGVYLAENLFLVQAPKSGDAQD--LQFYDKCLPGNSTVLNE- 1159
QY 1254 LGHRPAPVAALPPCEBLEGGLYMPQELTVSSVLVE-----LTDIVHCR 1300
Db 1160 -----FDVAVTMCSDI-----SLWVKCVLDFSKSVPLPRDNTKVPMTPI--R 1201
QY 1301 MAAPSOR-KAVLSTING-----RYGRRTKLYEAH 1329
Db 1202 TAERPRPSQGLLENLVAMIKRNFNSPELSTVDMENTASVADRFPDSTYFLKDKLSGCSL 1261
QY 1330 SDV-----RESLARI-----PTIGPVQATTCELYELVE-----AMVEKGODGSAY 1370
Db 1262 GDSGKNIIDROALIRMEKQEKSTIQGLAD-----YDFVLDLPAIDQYRHIKQPKQ-- 1314
QY 1371 LEIDLONRDVSRITFFOKXCNKFTTGFTLAHGKVGQGISAMSKTECALFGPWFRAIEKEI 1430
Db 1315 -KIDLSIQ-----SEYPSLQTIY-----HSKRINALFGPIFSELTROM 1352
QY 1431 LALLPNN--TFYGDAAVEESY--FAAAVSGAGSCVWFENDESEPTQNNFSLGLECYVME 1486
Db 1353 LSAIDTSRYLFTFRKTPQEJEEFFSDLAHQPMHEVLELDVSKIDKSQNEFHCAVEYEIWK 1412
QY 1487 ECGMPQWLRLVHLVRSAMTLQAPKESLKGf-----WKHSGEPGTLMLNTVWMA 1537
Db 1413 RLGIDEFILAV-----WKQGHRTKTLKDYTAGIKTKLWYQRKSGDVTFGTGVTIIA 1465
QY 1538 IIAHCYEFRRFRAAFKGDSDSVYLCSDYRQSRNAALIAAGCL-----KLKV 1584
Db 1466 CMASMLPMEKVIKAAFCGDDSLVY-----LPKCELPNIOCANLMMNFEAKL 1513
QY 1585 DYRPIGLYAGVVA--PGGTLPLDVYRFAGRLSEKNWGPGERAEQRLAVCDFLRLGLT 1641

```

```

Db 1514 FKTYGYFCGRVYIHHDRGAIYVYDPLKIISKLGAKHI-TDKEHLEEFRLSLADVSKSLN 1572
QY 1642 NVAQVCY--DVSRYGVGSP 1659
Db 1573 NCAYYAQDDEAVREVHKTAP 1592

```

Search completed: March 7, 2002, 14:08:31
Job time: 279 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:03:02 ; Search time 49.94 Seconds
(without alignments)
4973.376 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013
Sequence: 1 PGITTAIEQALAAANSALA.....FTETIKPVLDITNSIIOKRV 1698

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505.seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

SPTREMBL.17:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp rodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9001	99.9	1698	12	O9YLK3 hepatitis e
2	8812.5	97.8	1708	12	O9YK10 swine hepat
3	8780.5	97.4	1708	12	O9YLK1 hepatitis e
4	7730	85.8	1707	12	O9YLV29 hepatitis e
5	7475	82.9	1685	12	O9E8G6 hepatitis e
6	7468	82.9	1693	12	O69410 hepatitis e
7	7457	82.7	1693	12	O9W1L5 hepatitis e
8	7456	82.7	1693	12	O89444 hepatitis e
9	7424	82.4	1693	12	O81876 hepatitis e
10	7412	82.2	1693	12	O81862 hepatitis e
11	7333	81.4	1693	12	O9WC28 hepatitis e
12	7329	81.3	1693	12	O39221 hepatitis e
13	7304	81.0	1693	12	O81344 hepatitis e
14	7301	81.0	1693	12	O69418 hepatitis e
15	3444	38.2	727	12	O81873 hepatitis e
16	2520	28.0	479	12	O71144 hepatitis e
17	2365	26.2	452	12	O71145 hepatitis e
18	2284	25.3	487	12	O81869 hepatitis e
19	1821	20.2	491	12	O9YWL2 hepatitis e

20	1368	15.2	290	12	O81875 hepatitis e
21	1183	13.1	245	12	O81868 hepatitis e
22	1081	12.0	225	12	O9WLK5 hepatitis e
23	1074	11.9	217	12	O81864 hepatitis e
24	1025	11.4	212	12	O9YWL1 hepatitis e
25	988	11.0	210	12	O92745 hepatitis e
26	988	11.0	210	12	O92746 hepatitis e
27	988	11.0	210	12	O92747 hepatitis e
28	988	11.0	210	12	O92748 hepatitis e
29	988	11.0	210	12	O92749 hepatitis e
30	981	10.9	210	12	O56046 hepatitis e
31	913	10.1	193	12	O9WLK6 hepatitis e
32	884	9.8	182	12	O81863 hepatitis e
33	769	8.5	158	12	O9W9G2 hepatitis e
34	763	8.5	152	12	O81874 hepatitis e
35	709	7.9	134	12	O911J4 swine hepat
36	692	7.7	158	12	O81867 hepatitis e
37	666	7.4	134	12	O99HP9 hepatitis e
38	666	7.4	136	12	O9W8P3 hepatitis e
39	665	7.4	136	12	O9WR71 hepatitis e
40	662	7.3	133	12	O99HP1 hepatitis e
41	661	7.3	133	12	O99HP4 hepatitis e
42	658	7.3	136	12	O9W9Y6 hepatitis e
43	654	7.3	136	12	O9WR68 hepatitis e
44	646	7.2	123	12	O9E3S9 hepatitis e
45	646	7.2	123	12	O9DQAV hepatitis e

ALIGNMENTS

RESULT 1
O9YLK3 PRELIMINARY; PRT; 1698 AA.
AC O9YLK3;
DT 01-MAY-1999 (TRENBLER, 10, Created)
DT 01-MAY-1999 (TRENBLER, 10, Last sequence update)
DT 01-JUN-2001 (TRENBLER, 17, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
GN ORF1.
OS Hepatitis E virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HEV-US1;
RX MEDLINE=98178637; PubMed=9519822;
RA Schlauder G.G., Dawson G.J., Erker J.C., Kwo P.Y., Knigge M.F., Smalley D.L., Rosenblatt J.E., Desai S.M., Mushahwar I.K.;
RT "The sequence and phylogenetic analysis of a novel hepatitis E virus isolated from a patient with acute hepatitis reported in the United States.";
RT J. Gen. Virol. 79:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HEV-US1;
RX MEDLINE=99190429; PubMed=10092008;
RA Erker J.C., Desai S.M., Schlauder G.G., Dawson G.J., Mushahwar I.K.;
RT "A hepatitis E virus variant from the United States: molecular characterization and transmission in cynomolgus macaques.";
RT J. Gen. Virol. 80:681-690(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-HEV-US1;
RX Schlauder G.G., Erker J.C., Dawson G.J., Desai S.M., Knigge M.F., Kwo P.Y., Smalley D.L., Rosenblatt J.E., Mushahwar I.K.;
RT Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-HEV-US1;
RX Erker J.C., Schlauder G.G., Dawson G.J., Desai S.M., Mushahwar I.K.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF060668; AAD15812.1; -.

DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
DR Pfam: PF01661; DUF27; 1.
DR SMART: SM00306; A1pp; 1.
DR Polyprotein.
FT NON_TER
SQ SEQUENCE 1698 AA; 186069 MW; 1C670601150F4EA CRC64;

Query Match 99.9%; Score 9001; DB 12; Length 1698;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCITTAIEQAAALAAANSALANAVVPRFLSRVOTELINLMOPROLVPRPEVLMMHPRIOR 60
DB 1 PCITTAIEQAAALAAANSALANAVVPRFLSRVOTELINLMOPROLVPRPEVLMMHPRIOR 60
QY 61 VINHELEQYCARAGRCLEVGAAHPSINDPNVLRHCELRPGRDVQRYSAPTGPAAN 120
DB 61 VINHELEQYCARAGRCLEVGAAHPSINDPNVLRHCELRPGRDVQRYSAPTGPAAN 120
QY 121 CRRSALRGLPPADRTYCEDEGFSRCAFAETGVALYSLHDLMPADVAEAMARHGXTRYAA 180
DB 121 CRRSALRGLPPADRTYCEDEGFSRCAFAETGVALYSLHDLMPADVAEAMARHGXTRYAA 180
QY 181 LHLPREVLLPCTGHTTSTYLLHGDGRANVYEGDTSAGYNNDVSTLRWITTKIYVGH 240
DB 181 LHLPREVLLPCTGHTTSTYLLHGDGRANVYEGDTSAGYNNDVSTLRWITTKIYVGH 240
QY 241 PLVIERVNAIGHFVLLTAAPESPMPVYPRSTEVVRSIFGPGSPLFPSACSTR 300
DB 241 PLVIERVNAIGHFVLLTAAPESPMPVYPRSTEVVRSIFGPGSPLFPSACSTR 300
QY 301 STEHAVPYHIMDRMLFGATLDDQAFCCSRMLTYLRGISYKVTGALVANEGMNASDAL 360
DB 301 STEHAVPYHIMDRMLFGATLDDQAFCCSRMLTYLRGISYKVTGALVANEGMNASDAL 360
QY 361 TAXITAAVLTICHOVYLRFOATISKGMRLGVENHAKFTIRLXSNLFEKSGRXYTGROQ 420
DB 361 TAXITAAVLTICHOVYLRFOATISKGMRLGVENHAKFTIRLXSNLFEKSGRXYTGROQ 420
QY 421 FYAOCRRMLASAGFHLDPRLVFEDESVPCRTFLKVKAGKFCFCFMRMLGOECTCELEPAE 480
DB 421 FYAOCRRMLASAGFHLDPRLVFEDESVPCRTFLKVKAGKFCFCFMRMLGOECTCELEPAE 480
QY 481 GLVGDHGDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNVPDITAAASRLT 540
DB 481 GLVGDHGDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNVPDITAAASRLT 540
QY 541 ATVELVASPDRLRECTVGNKTEFTTVYDGAHLEANGPROYLSLSDASROSGAGSHSLT 600
DB 541 ATVELVASPDRLRECTVGNKTEFTTVYDGAHLEANGPROYLSLSDASROSGAGSHSLT 600
QY 601 YELTPAGLOVRISSNGLDCTATFPPGASAPSAPEGVAACFCSALYRYNRTQHSITLGLIM 660
DB 601 YELTPAGLOVRISSNGLDCTATFPPGASAPSAPEGVAACFCSALYRYNRTQHSITLGLIM 660
QY 661 LHPBGLGIFPPSPGHTWESANPFCGEGTLYTRTWSTSGESSDFSPREAAAPAMAATPG 720
DB 661 LHPBGLGIFPPSPGHTWESANPFCGEGTLYTRTWSTSGESSDFSPREAAAPAMAATPG 720
QY 721 LPHSTPPVSDIHWLPSPSEEROVDAAPVAPADPAGLPGPVLLTPPPPVAKPIPPPS 780
DB 721 LPHSTPPVSDIHWLPSPSEEROVDAAPVAPADPAGLPGPVLLTPPPPVAKPIPPPS 780
QY 781 RNRRLLYTPDGAKYVAGSLFESDCDWLVNASNPGHRPGGGLCHAFYORFPAEFPTFEI 840
DB 781 RNRRLLYTPDGAKYVAGSLFESDCDWLVNASNPGHRPGGGLCHAFYORFPAEFPTFEI 840
QY 841 MREGIAATTLPRPIIHAAPDYRVEQNPKRLEAAVRETCSRGTAAVPLLGSGIYQVPV 900

DB 841 MREGIAATTLPRPIIHAAPDYRVEQNPKRLEAAVRETCSRGTAAVPLLGSGIYQVPV 900
QY 901 SUSPAMERNHRPGDELYLTPPAAMFPAOPPLTTEPTARTANLALIDAATEVG 960
DB 901 SUSPAMERNHRPGDELYLTPPAAMFPAOPPLTTEPTARTANLALIDAATEVG 960
QY 961 RACAGCTISPGIVHYOFTAGVPSGSKSRISIOGDVDDVVVPTRELNSMRRGFAAFTH 1020
DB 961 RACAGCTISPGIVHYOFTAGVPSGSKSRISIOGDVDDVVVPTRELNSMRRGFAAFTH 1020
QY 1021 TAARVTIGRRVVIDAPSLPRLLLHMQRASSYHLGDPNQIPALIDENHAGLVPAIRE 1080
DB 1021 TAARVTIGRRVVIDAPSLPRLLLHMQRASSYHLGDPNQIPALIDENHAGLVPAIRE 1080
QY 1081 LAPTSMMXYTHRCPADVCLELIGAYPKIOTTSRVLSLEWNPAPALGOKLVYXQAANKAP 1140
DB 1081 LAPTSMMXYTHRCPADVCLELIGAYPKIOTTSRVLSLEWNPAPALGOKLVYXQAANKAP 1140
QY 1141 GAITVHEAGATFETTTIATADARGLIOSSRAHAIVALTRTEKCVIILDARGLLEVG 1200
DB 1141 GAITVHEAGATFETTTIATADARGLIOSSRAHAIVALTRTEKCVIILDARGLLEVG 1200
QY 1201 SDVIYNNFTLAGGEVGHRSYIIPGNPDQNLGTLQAPPSCQISAYHQLABELGHRAP 1260
DB 1201 SDVIYNNFTLAGGEVGHRSYIIPGNPDQNLGTLQAPPSCQISAYHQLABELGHRAP 1260
QY 1261 VAAVLPCPELEOGLLYMPOELTVSDSYLVFELTDIVHCRMAAPSORKAVLSTLVGRYGR 1320
DB 1261 VAAVLPCPELEOGLLYMPOELTVSDSYLVFELTDIVHCRMAAPSORKAVLSTLVGRYGR 1320
QY 1321 RTKLYEAASHDVRESLARIPTIGPVQATTCLELYELVENAMVERKODGSAYLELDICNRDV 1380
DB 1321 RTKLYEAASHDVRESLARIPTIGPVQATTCLELYELVENAMVERKODGSAYLELDICNRDV 1380
QY 1381 SRTTFQXCKNFTTGETIARHKVQOGISAMSKTCALFGPMFRAIERETLLALRPNTFY 1440
DB 1381 SRTTFQXCKNFTTGETIARHKVQOGISAMSKTCALFGPMFRAIERETLLALRPNTFY 1440
QY 1441 GDAYEESYFAAASVAGSGCMVFENDFSEPTQNNFSLGLECVMBEGCMQMLIRLYHL 1500
DB 1441 GDAYEESYFAAASVAGSGCMVFENDFSEPTQNNFSLGLECVMBEGCMQMLIRLYHL 1500
QY 1501 VASAMILQAPKESLKGFMKKHSGEBGTLMTVMMAITIAHCYERDFRVAAFKGGDSYV 1560
DB 1501 VASAMILQAPKESLKGFMKKHSGEBGTLMTVMMAITIAHCYERDFRVAAFKGGDSYV 1560
QY 1561 LCSIDYRQSRNAALLAGCGGLKLVDRPRTGLYAGVVAAGLGLDLYVAFAGRLSEKMMG 1620
DB 1561 LCSIDYRQSRNAALLAGCGGLKLVDRPRTGLYAGVVAAGLGLDLYVAFAGRLSEKMMG 1620
QY 1621 PGPERAEOLRLAVCDFLGLTVAOVCDVYGVSRVYSGVLVHNLIGMLQTLADGKAHPT 1680
DB 1621 PGPERAEOLRLAVCDFLGLTVAOVCDVYGVSRVYSGVLVHNLIGMLQTLADGKAHPT 1680
QY 1681 ETIKPVLDTNSIIQORVE 1698
DB 1681 ETIKPVLDTNSIIQORVE 1698

RESULT 2
Q9YK10 PRELIMINARY; PRT; 1708 AA.
AC Q9YK10:
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE.
OS Swine hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=63421.
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=MENG;
RA MEDLINE-97420774; PubMed-9275216;
RX Meng X.J., Purcell R.H., Halbur P.G., Lehman J.R., Webb D.M.,
RA Tsareva T.S., Haynes J.S., Thacker B.J., Emerson S.U.;
RT "A novel virus in swine is closely related to the human hepatitis E
RT virus";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9860-9865(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MENG;
RX MEDLINE-99030877; PubMed-9811705;
RA Meng X.J., Halbur P.G., Shapiro M.S., Govindarajan S., Bruna J.D.,
RA Mushahar I.K., Purcell R.H., Emerson S.U.;
RT "Genetic and experimental evidence for cross-species infection by
RT swine hepatitis E virus";
RL J. Virol. 72:9714-9721(1998).
DR EMBL: AF082843; AAC97208.1; -;
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; Viral_helicase1.
DR InterPro: IPR002589; DUF27.
DR pfam: PF01443; Viral_helicase1; 1.
DR pfam: PF01660; Vmethyltransf; 1.
DR pfam: PF01661; DUF27; 1.
DR SMART: SM00506; A1pp; 1.
KW Polypeptide.
FT CHAIN 55 237 METHYLTRANSFERASE.
FT CHAIN 975 1219 HELICASE.
FT CHAIN 1222 1708 RNA-DIRECTED RNA POLYMERASE.
SQ SEQUENCE 1708 AA; 187406 MW; 7444E52DCD616130 CRC64;

Query Match 97.8%; Score 8812.5; DB 12; Length 1708;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1661; Conservative 7; Mismatches 30; Indels 1; Gaps 1;

QY 1 PGITTAIEQAALAAANASALANAVVPRFLSRVOTELINIMQPROLVFREVLMNHP10R 60
DB 10 PGITTAIEQAALAAANASALANAVVPRFLSRVOTELINIMQPROLVFREVLMNHP10R 69
QY 61 VHNLEIYQCRARAGCLEYGAHPRISINDPNVLRHCFLEPRVGRDQWRYSAPTRGPAAN 120
DB 70 AHNLEIYQCRARAGCLEYGAHPRISINDPNVLRHCFLEPRVGRDQWRYSAPTRGPAAN 129
QY 121 CRRSALRGIPADRTYCFGFSRCAPAEETGVALYSLHDLMPADVAEAMARHGXTRLYAA 180
DB 130 CRRSALRGIPADRTYCFGFSRCAPAEETGVALYSLHDLMPADVAEAMARHGXTRLYAA 189
QY 181 LHLPEVLLPFGTYHTTSYLLIHDGRAVYTYEGDSAGYNHDVSLIRAMIRTTKIYGDH 240
DB 190 LHLPEVLLPFGTYHTTSYLLIHDGRAVYTYEGDSAGYNHDVSLIRAMIRTTKIYGDH 249
QY 241 PLVIEVRALIGCHFVLLTAAPESPMPVYPRSTEVYRSIFGPGSPSLPPSACSTK 300
DB 250 PLVIEVRALIGCHFVLLTAAPESPMPVYPRSTEVYRSIFGPGSPSLPPSACSTK 309
QY 301 STFNHAYVNHMDLMPGATLDQAFCSSRLMTYLGISIKYVYVGAALVANEGMNASDAL 360
DB 310 STFNHAYVNHMDLMPGATLDQAFCSSRLMTYLGISIKYVYVGAALVANEGMNASDAL 369
QY 361 TAYITAAVLTICHQRLTRQAIKSGMRRLGVEHAOKFTTLYSMLPEKSGRDYIPGRQLO 420
DB 370 TAYITAAVLTICHQRLTRQAIKSGMRRLGVEHAOKFTTLYSMLPEKSGRDYIPGRQLO 429
QY 421 FYAQCRRMLISAGFHLDPRVLVPEDESVPICRRTFLKVKVAGFCFMRWLGQECTCLEPAE 480
DB 430 FYAQCRRMLISAGFHLDPRVLVPEDESVPICRRTFLKVKVAGFCFMRWLGQECTCLEPAE 489
QY 481 GLVGDGHNDNEAYEGSEVDPAPBAHLDVSTYAVHGHQLEFALYRALNVPDIIAARSRLT 540
DB 490 GLVGDGHNDNEAYEGSEVDPAPBAHLDVSTYAVHGHQLEFALYRALNVPDIIAARSRLT 549
QY 541 ATVELVASPRLRCRTVLGKRTFTTYVDGAHLFANGPDEQYVLSFDSRQSMGAGSHSLT 600

DB 550 ATVELVASPRLRCRTVLGKRTFTTYVDGAHLFANGPDEQYVLSFDSRQSMGAGSHSLT 609
QY 601 YELTPAGLOVRISNSGLDCAATPPPGAPSAABGEVAAPFCASALYRNRFORSLTGLW 660
DB 610 YELTPAGLOVRISNSGLDCAATPPPGAPSAABGEVAAPFCASALYRNRFORSLTGLW 669
QY 661 IMPEGLIGTFPPSPGHIMESANPCGEGTLYRTSTSGSSDSPPEAAAPAMATPG 720
DB 670 IMPEGLIGTFPPSPGHIMESANPCGEGTLYRTSTSGSSDSPPEAAAPAMATPG 729
QY 721 LPHSTPVSIDWLPPPESEFOVDAAPVPAPDAGLPGVVLTP-PPPEPVHKPSITPP 779
DB 730 LPHSTPVSIDWLPPPESEFOVDAAPVPAPDAGLPGVVLTP-PPPEPVHKPSITPP 789
QY 780 SRNRRLTYTPDCAKYVAGSLFESSDDMLVNASNPNHRPGGGLCHAFYQFPAFAFPTER 839
DB 790 SRNRRLTYTPDCAKYVAGSLFESSDDMLVNASNPNHRPGGGLCHAFYQFPAFAFPTER 849
QY 840 IMREGLAAYTLTPRPIITHAAPDYRYEQNPKRLEAAVRETCSSRGTAAYPLGSGIYQV 899
DB 850 IMREGLAAYTLTPRPIITHAAPDYRYEQNPKRLEAAVRETCSSRGTAAYPLGSGIYQV 909
QY 900 VLSLFDAMERNHRPDELVLTPEPAAMFEANKPAQPALTTTEDTARTANLALTEIDATEV 939
DB 910 VLSLFDAMERNHRPDELVLTPEPAAMFEANKPAQPALTTTEDTARTANLALTEIDATEV 969
QY 960 GRACAGCTISPGIVHQTFTAGVSGSKRSIQGGDVVVVYPTRELNSRRRGFAFTP 1019
DB 970 GRACAGCTISPGIVHQTFTAGVSGSKRSIQGGDVVVVYPTRELNSRRRGFAFTP 1029
QY 1020 HTAARVTIGRRVYIDEAPSLPHLLLMORASSVHLGDPNOIPATIDFAGLVPATRP 1079
DB 1030 HTAARVTIGRRVYIDEAPSLPHLLLMORASSVHLGDPNOIPATIDFAGLVPATRP 1089
QY 1080 ELAPTSWMVYTHRCPADVCELIRGAYPKIQTTSRVLSLFWNEPAIGOKLVYTOAAKAA 1139
DB 1090 ELAPTSWMVYTHRCPADVCELIRGAYPKIQTTSRVLSLFWNEPAIGOKLVYTOAAKAA 1149
QY 1140 PGATVYHAGAGFTETTTATADARGLIOSSRAHAVALTRHREKVILIDAGGLREVG 1199
DB 1150 PGATVYHAGAGFTETTTATADARGLIOSSRAHAVALTRHREKVILIDAGGLREVG 1209
QY 1200 ISDVIYNNFVLAGEVGHXRPVYPRGNPDONICTLOAFPPSCIOISYHOLAEBELGRPA 1259
DB 1210 ISDVIYNNFVLAGEVGHXRPVYPRGNPDONICTLOAFPPSCIOISYHOLAEBELGRPA 1269
QY 1260 PVAAYLPPCELEBOGLYMPQELTVSDVLFELTDLVHCRMAAPSORKAVLSTLYGRYG 1319
DB 1270 PVAAYLPPCELEBOGLYMPQELTVSDVLFELTDLVHCRMAAPSORKAVLSTLYGRYG 1329
QY 1320 RRTKLYEAAHSDVRESIARFTPTIGVQVATTCELVELVEAMVERKGDGSAVLELDLCNRD 1379
DB 1330 RRTKLYEAAHSDVRESIARFTPTIGVQVATTCELVELVEAMVERKGDGSAVLELDLCNRD 1389
QY 1380 VSRITTFQKCKNKFPTGETTIAHGKVGOGISAMSTFCALGPMFRALTEKETLILPNNIF 1439
DB 1390 VSRITTFQKCKNKFPTGETTIAHGKVGOGISAMSTFCALGPMFRALTEKETLILPNNIF 1449
QY 1440 YGDAYEESVFAAASGASCMVFENDSEFDSTONNLSLGLCYVMECCGMQWILILYH 1499
DB 1450 YGDAYEESVFAAASGASCMVFENDSEFDSTONNLSLGLCYVMECCGMQWILILYH 1509
QY 1500 LVRSAMITLQAPKESLKGFMKKHSGEPSTLLMNTVWNAITAHCEYFDFRVAAEKGDSDV 1559
DB 1510 LVRSAMITLQAPKESLKGFMKKHSGEPSTLLMNTVWNAITAHCEYFDFRVAAEKGDSDV 1569
QY 1560 VLSGDYRQSRNAALILAGCGIKLVDRPRTGLVAGVVAAGGLTLPVYVFAAGLSKRNW 1619
DB 1570 VLSGDYRQSRNAALILAGCGIKLVDRPRTGLVAGVVAAGGLTLPVYVFAAGLSKRNW 1629
QY 1620 GPGPERAEOLRLAVCDFRLGLTJNVAQCVYVSVYVSGPLVNLIGMLOTINDGKAHF 1679
DB 1630 GPGPERAEOLRLAVCDFRLGLTJNVAQCVYVSVYVSGPLVNLIGMLOTINDGKAHF 1689

```
QY 1680 TETIKPVLDTNLSIQORVE 1698
      |||||||
Db 1690 TETIKPVLDTNLSIIORVE 1708

RESULT 3
QYLR1
ID QYLR1 PRELIMINARY: PRT: 1708 AA.
AC QYLR1:
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE POLYPEPTIDEIN.
GN ORF1.
OC Hepatitis E virus.
OS Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US2;
RX MEDLINE=99178637; PubMed=9519822;
RA Schauder G.G., Dawson G.J., Erker J.C., Kwo P.Y., Knigge M.F.,
RA Smalley D.L., Rosenblatt J.E., Desai S.M., Mushahwar I.K.;
RT "The sequence and phylogenetic analysis of a novel hepatitis E virus
RT isolated from a patient with acute hepatitis reported in the United
RT States.";
RL J. Gen. Virol. 79:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US2;
RX MEDLINE=99190429; PubMed=10092008;
RA Erker J.C., Desai S.M., Schauder G.G., Dawson G.J., Mushahwar I.K.;
RA "A hepatitis E virus variant from the United States: molecular
RT characterization and transmission in cynomolgus macaques.";
RL J. Gen. Virol. 80:681-690(1999).
DR EMBL: AF060669; AAD15815.1; -
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1.
DR Pfam: PF01660; Viral_helicase1.
DR Pfam: PF01661; DUF27; 1.
DR SMART: SM00506; A1pp; 1.
DR Polyprotein.
SQ SEQUENCE 1708 AA; 187260 MW; BE0DF5EEEDAB547 CRC64;

Query Match 97.4%; Score 8780.5; DB 12; Length 1708;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1655; Conservative 8; Mismatches 35; Indels 1; Gaps 1;
```

```
Db 310 SFHAAPVHIMDXLMFGATLXDOAFCCSRIMTYLIGISYKVTGALVANEGMNASDAL 369
      |||||||
QY 361 TAXITAAVITLICHORLRTQALISKGRRLGVEHAQCFITRLYSMLFEKSGROYIPROQ 420
      |||||||
Db 370 TAVITAAVLTICHORLRTQALISKGRRLGVEHAQCFITRLYSMLFEKSGROYIPROQ 429
      |||||||
QY 421 FYAOCRRMLASGFIHDPRLVFEDESVPCRCRTFLKVKAGFCFCFMMLCOECTCFLEPAE 480
      |||||||
Db 430 FYAOCRRMLASGFIHDPRLVFEDESVPCRCRTFLKVKAGFCFCFMMLCOECTCFLEPAE 489
      |||||||
QY 481 GLVGDHGDHNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNPQDIAAARSRLT 540
      |||||||
Db 490 GLVGDHGDHNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNPQDIAAARSRLT 549
      |||||||
QY 541 ATVELVASPDRLRCRTVLCNKTFRITVDGAILHANGPQOYVLSPASQSGAGSHSLT 600
      |||||||
Db 550 ATVELVASPDRLRCRTVLCNKTFRITVDGAILHANGPQOYVLSPASQSGAGSHSLT 609
      |||||||
QY 601 YELTPAGLQVRISNGLDCTATFPFGAPSAAPGEVAACFSLYRNRTORHSLTGLM 660
      |||||||
Db 610 YELTPAGLQVKTSSNGLDCTATFPFGAPSAAPGEVAACFSLYRNRTORHSLTGLM 669
      |||||||
QY 661 LHPEGLIGIFPPSPGHIMESANPFCGECTLYRTWSTSGFSDESPPPAAPAMAAATPG 720
      |||||||
Db 670 LHPEGLIGIFPPSPGHIMESANPFCGECTLYRTWSTSGFSDESPPPAAPAMAAATPG 729
      |||||||
QY 721 LPHSTPPVSDIVLPPSPSEFOYDAAPVPADPDAGLPEPVLT -PPPPPVHKPSITPP 779
      |||||||
Db 730 LPPVPPVSDIVLPPSPSESHVDAASVPSPVEPGLTSPVLTTPPPPPVPRKAPTSPP 789
      |||||||
QY 780 SNRRRLTYTPDGAKYAGSLFESDCDMLVMSNPGHRGGGLCHAFYRPEAPYPTPF 839
      |||||||
Db 790 PSTRRLTYTPDGAKYAGSLFESDCDMLVMSNPGHRGGGLCHAFYRPEAPYPTPF 849
      |||||||
QY 840 IIRREGIAATLTPRPIIAHAPADYRVEONPKRLAAAYRETCSRKGTAAYPLLGSIYOVP 899
      |||||||
Db 850 IIRREGIAATLTPRPIIAHAPADYRVEONPKRLAAAYRETCSRKGTAAYPLLGSIYOVP 909
      |||||||
QY 900 VLSFDAMERNRHPDDELTLTPPAANMFANPAPQVLTITDTPRTATLALIDEAATPV 959
      |||||||
Db 910 VLSFDAMERNRHPDDELTLTPPAANMFANPAPQVLTITDTPRTATLALIDEAATPV 969
      |||||||
QY 960 GRACAGCTISPGIVAYOFTAGVPGSGKRSIQGQDVVVVPTRELNSWRRRGFAATFP 1019
      |||||||
Db 970 GRACAGCTISPGIVAYOFTAGVPGSGKRSIQGQDVVVVPTRELNSWRRRGFAATFP 1029
      |||||||
QY 1020 HTAAVYTIIGRRVYIDEASLPRLHLLLMQARSSVHLGDPQIPALIDEHAGLVPAIRP 1079
      |||||||
Db 1030 HTAAVYTIIGRRVYIDEASLPRLHLLLMQARSSVHLGDPQIPALIDEHAGLVPAIRP 1089
      |||||||
QY 1080 ELAPTSMXVTHRCPADVCELRGAYPKIQTSTRVLSLFENNEPAILGOKLVYTOAKKAN 1139
      |||||||
Db 1090 ELAPTSMXVTHRCPADVCELRGAYPKIQTSTRVLSLFENNEPAILGOKLVYTOAKKAN 1149
      |||||||
QY 1140 PGATVHEAGATFTETTLIATADARGLIOSSRAHAIVALTHTKCYVLIDAPGLREVG 1199
      |||||||
Db 1150 PGATVHEAGATFTETTLIATADARGLIOSSRAHAIVALTHTKCYVLIDAPGLREVG 1209
      |||||||
QY 1200 ISDVIVNNEFLAGGEVGAHRSVITRGNPDONLGLQAFPPSCQISATHQLAEBELGHRPA 1259
      |||||||
Db 1210 ISDVIVNNEFLAGGEVGAHRSVITRGNPDONLGLQAFPPSCQISATHQLAEBELGHRPA 1269
      |||||||
QY 1260 PVAALVPCPPELQGLLVMPOELTYSDSVLYVELLDIYHCRMAAPSOKRAVLTIVGRG 1319
      |||||||
Db 1270 PVAALVPCPPELQGLLVMPOELTYSDSVLYVELLDIYHCRMAAPSOKRAVLTIVGRG 1329
      |||||||
QY 1320 RRTKLYEAHSDVRESLARFTPTIGPVQATTCETELVEAMVEKQODSAYVELDLQCNBD 1379
      |||||||
Db 1330 RRTKLYEAHSDVRESLARFTPTIGPVQATTCETELVEAMVEKQODSAYVELDLQCNBD 1389
      |||||||
QY 1380 VSRITFFQKXCKFTTGETTIAHGKVGQGISAMSKTFCALFGWPAIRIEKEITALLPNI 1439
      |||||||
```

Db 1390 VSRITFEQKCNKFTTGETTIAHGKVGQGISAMSKTECALFGPWFRAIEKEILLAPENIF 1449
QY 1440 YGSAVESVFAAASGAGSGMVEENDESEEDSONNPSISGLECVMEECGMQWLTRLYH 1499
Db 1450 YGSAVESVFAAASGAGSGMVEENDESEEDSONNPSISGLECVMEECGMQWLTRLYH 1509
QY 1500 LVMSAMTLOAPKESLKGFMKHSGEPTLLMNTVMNMAIIAHCEYEFDFVAAAFKGDSDV 1559
Db 1510 LVMSAMTLOAPKESLKGFMKHSGEPTLLMNTVMNMAIIAHCEYEFDFVAAAFKGDSDV 1569
QY 1560 VLCSDRQSRNMAALAGCGLKLVDRPILGYAGVVAADGLTLPDVVPAGRLSEKNV 1619
Db 1570 VLCSDRQSRNMAALAGCGLKLVDRPILGYAGVVAADGLTLPDVVPAGRLSEKNV 1629
QY 1620 GGPBERAEQRLAVCDRLTNVAVQCVVSVRYGVSPGLVHNLIGMQLTADGAHF 1679
Db 1630 GGPBERAEQRLAVCDRLTNVAVQCVVSVRYGVSPGLVHNLIGMQLTADGAHF 1689
QY 1680 TETIKPVLDTNSIIORVE 1698
Db 1690 TENIKPVLDTNSIIORVE 1708
RESULT 4
QY1V29 PRELIMINARY; PRT: 1707 AA.
AC QY1V29;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE POLYPROTEIN.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
ON NCBL_TaxID=12461;
OX [1]
RN RC STRAIN=tl;
RP SEQUENCE FROM N.A.
RA Harrison T.J.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=tl;
RX MEDLINE=20318653; PubMed=10859372;
RA Wang Y., Zhang H., Ling R., Li H., Harrison T.J.;
RT "The complete sequence of hepatitis E virus genotype 4 reveals an
RT alternative strategy for translation of open reading frames 2 and 3.";
RL J. Gen. Virol. 81:1675-1686(2000).
DR EMBL: AJ272108; CAB83209.1; -;
DR InterPro: IPR002589; DUF27.
DR InterPro: IPR000606; Viral_helicase.
DR InterPro: IPR002388; V_methyltransf.
DR Pfam: PF01661; DUF27.1.
DR Pfam: PF01443; Viral_helicase1.1.
DR Pfam: PF01660; Vmethyltransf.1.
DR SMART: SM00506; A1pp.1.
SO SEQUENCE 1707 AA, 187296 MW, 2A80F303AB96C4EC CRC64;

Query Match 85.8%; Score 7730; DB 12; Length 1707;
Best local similarity 83.9%; Pred. No. 0;
Matches 1444; Conservative 91; Mismatches 140; Indels 46; Gaps 5;

QY 181 LHLPEVLLPGTYHTSYLLIHGDRAVVTYEDGTSGAGNHVSLRAMVIRTKIYGDH 240
Db 190 LHLPEVLLPGTYHTSYLLIHGDRAVVTYEDGTSGAGNHVSLRAMVIRTKIYGDH 249
QY 241 PLVIERVRAIGCHVLLTLAEPSPMPVYPVPRSTEVYRSIFPGGSPSLFSPASCTK 300
Db 250 PLVIERVRAIGCHVLLTLAEPSPMPVYPVPRSTEVYRSIFPGGSPSLFSPASCTK 309
QY 301 STHNAVPHIWDRLMFLGALDDOAFCCSKLMTYLRKISIKYVYGALVANEGNASDAL 360
Db 310 STHNAVPHIWDRLMFLGALDDOAFCCSKLMTYLRKISIKYVYGALVANEGNASDAL 369
QY 361 TAYITAYLTTCORRYLRTQAIKGMRLVLEHAKKITLYLSWLFKSGRDYIPGQLO 420
Db 370 TAYITAYLTTCORRYLRTQAIKGMRLVLEHAKKITLYLSWLFKSGRDYIPGQLO 429
QY 421 FYAQRRLWSAGFHLDRVLVFEDEVPICRGTFLKKAAGFCCPMRIAGDCTCFLEPAE 480
Db 430 FYAQRRLWSAGFHLDRVLVFEDEVPICRGTFLKKAAGFCCPMRIAGDCTCFLEPAE 489
QY 481 GLVGDHNDNAYEGSEVDPAEPAHLVDYSGTYAVHGHOLEALYRALVPODIAARSLT 540
Db 490 GRVGEQGYDNEAEFGSDIDPAEPAVSIAGSYIVTGSQLOPLYOALGIPSDLAARSLT 549
QY 541 ATVEIVASPDRLBCRYTLGKKTFTTYVDGAHLEANGPEQVILSEFDSRMSGSHSLT 600
Db 550 ATVEIVASPDRLBCRYTLGKKTFTTYVDGAHLEANGPEQVILSEFDSRMSGSHSLT 609
QY 601 YELTPAGLQVIRISSNGIDCATPPPGAPSAPEVAFAFCALRYNRFTQRLHSLTGLW 660
Db 610 YELTPAGLQVIRISSNGIDCATPPPGAPSAPEVAFAFCALRYNRFTQRLHSLTGLW 669
QY 661 LHPBGLIGIFPPSPGHIWESANFCGEGTLYRTWTSGFSSDFSPPEAAAPMAATPG 720
Db 670 YHPBGLIGIFPPSPGHIWESANFCGEGTLYRTWTSGFSSDFSPPEAAAPMAATPG 724
QY 721 LPHSTPVSIDVLPPESE---EQVDAAP-----VPPAPDPAGLP----- 758
Db 725 LPHSTPVSIDVLPPESE---EQVDAAP-----VPPAPDPAGLP----- 770
QY 759 -GPVLTLPPLPPPPVYKPSIPPPSRNRRLTYPPGAKVYAGSLFESDCDMLVNASNGHR 817
Db 771 SGAPLAPPAALPYTHLSG---RRLLHYTPGSKVYAGSLFESDCTWLVNASNGHR 826
QY 818 PGGLCHAFYQRFPEAEYPTFEIMREGIAAYTLPRPIHAHADVRYEONPKRLEAYR 877
Db 827 PGGLCHAFYQRFPEAEYPTFEIMREGIAAYTLPRPIHAHADVRYEONPKRLEAYR 886
QY 878 ETCSRRTAAYPLDLSGTYOVPVSLSDAMERRNRPGDELTLTPPAANFPAKPAPVYL 937
Db 887 ETCSRRTAAYPLDLSGTYOVPVSLSDAMERRNRPGDELTLTPPAANFPAKPAPVYL 946
QY 938 TITEDTARTANLALEIDAATFVGACAGCTSPGIVHYOFTAGVGSKRSIOGDVDY 997
Db 947 TITEDTARTANLALEIDAATFVGACAGCTSPGIVHYOFTAGVGSKRSIOGDVDY 1006
QY 998 VVPTRELNSWRRRGFAAFTPHTAARYTIGRRVVIDEAPSLPPLLHLHMORASSVHL 1057
Db 1007 IVVPTRELNSWRRRGFAAFTPHTAARYTIGRRVVIDEAPSLPPLLHLHMORASSVHL 1066
QY 1058 GDPNQTALIDEHAGLVPAIRPELAPTSWXYVTHRCPADVCCELLRGAPKQTTSRYLRS 1117
Db 1067 GDPNQTALIDEHAGLVPAIRPELAPTSWXYVTHRCPADVCCELLRGAPKQTTSRYLRS 1126
QY 1118 LFWNEPAIGOKLVYTOAKANPAGAITVHEOGATFTETTTIATADAGLTOSSRAHIV 1177
Db 1127 LFWNEPAIGOKLVYTOAKANPAGAITVHEOGATFTETTTIATADAGLTOSSRAHIV 1186
QY 1178 ALTRHTEKCVVADPGLLREVGISDVIYVNFELAGGEVGHKRPVIVPRGNDQNLGTLQA 1237
Db 1187 ALTRHTEKCVVADPGLLREVGISDVIYVNFELAGGEVGHKRPVIVPRGNDQNLGTLQA 1246

QY 1238 FPPSCQISAVHQAELGHRPAVAALVPCPELEGGLLYMPQELTVSDVLFELTDIV 1297
DB 1247 FPPSCQISAVHQAELGHRPAVAALVPCPELEGGLLYMPQELTVSDVLFELTDIV 1306
QY 1298 HCMAPPSQKAVLSTLVNGYGRRTKLYEAHSDVRESLARFTPTGPGVATTCCELYEY 1357
DB 1307 HCMAPPSQKAVLSTLVNGYGRRTKLYEAHSDVRESLARFTPTGPGVATTCCELYEY 1366
QY 1358 EAMVEKGODSAVLELDLCNRDVSRTTFQKCNKFTTGETTIAHGKVGQISAMSKTFC 1417
DB 1367 EAMVEKGODSAVLELDLCNRDVSRTTFQKCNKFTTGETTIAHGKVGQISAMSKTFC 1426
QY 1418 LKSPMRALKEKELLALPPNIFYGAYEESVFAAVSAGSCWVFENDESEEDSTONNES 1477
DB 1427 LKSPMRALKEKELLALPPNIFYGAYEESVFAAVSAGSCWVFENDESEEDSTONNES 1486
QY 1478 LGLECVVMECCGPMOLIRLYHLYRSAMILLQAKESLKGPMKHSEBPGLLNTYWNNA 1537
DB 1487 LGLECVVMECCGPMOLIRLYHLYRSAMILLQAKESLKGPMKHSEBPGLLNTYWNNA 1546
QY 1538 IIAHCEFRDFAVAEFKGDSDVVLCSQDYRSRMAALLIAGCGKLKYDYKPIGLYAGVYV 1597
DB 1547 VIAHCYEFBDLKYAAFKGDSVVLCSQDYRSRMAALLIAGCGKLKYDYKPIGLYAGVYV 1606
QY 1598 AAGLGLTPVYVFRAGLSEKNMGCPGPERABQRLAYCDLRLGTLNVAQVCVDVYSRVGY 1657
DB 1607 AAGLGLTPVYVFRAGLSEKNMGCPGPERABQRLAYCDLRLGTLNVAQVCVDVYSRVGY 1666
QY 1658 SPGLVHNLIGMLQTLADGKAHFETIKPYLDLNTSTIIORE 1698
DB 1667 SPGLVHNLIGMLQTLADGKAHFETIKPYLDLNTSTIIORE 1707

RESULT 5
Q9E8G6 PRELIMINARY: PRT: 1685 AA.
ID Q9E8G6 AC Q9E8G6
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN (FRAGMENT).
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_taxid=12461;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=AB-2B.
RA van Cuyck-Gandre H., Zhang H.Y., Tsarev S.A., Warren R.L.,
RA Caudill J.D., Snellings N.O., Begot L., Innis B.L., Longet C.F.,
RA Tichhurst J.;
RA "Phylogenetically distinct hepatitis E viruses (HEC) in Pakistan";
RL Am. J. Trop. Med. Hyg. 0:0-0(1999).
DR EMBL: AF15822; AAG16764.1;
DR InterPro: IPR002589; DUF27.
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR Pfam: PF01661; DUF27; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
DR SMART: SM00506; Alpp; 1.
KW Polyprotein.
FT NON_TER 1
SO SEQUENCE 1685 AA; 184081 MW; 87CA74659104538B CRC64;

Query Match 82.9%; Score 7475; DB 12: Length 1685;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1400; Conservative 112; Mismatches 155; Indels 48; Gaps 6;

OY 1 PGITTAICQALAAANSALANAVVVRPFLSRVOTELLINMQPRLQVPRPEVLAMHPRIOR 60
DB 2 PGITTAICQALAAANSALANAVVVRPFLSRVOTELLINMQPRLQVPRPEVLAMHPRIOR 61

QY 61 VIHNELEQYCARAGCLEVGAHPRSINDPNVLRHCFLRPVGRDQYRWYSAPTRGPAAN 120
DB 62 VIHNELEQYCARAGCLEVGAHPRSINDPNVLRHCFLRPVGRDQYRWYSAPTRGPAAN 121
QY 121 CRSSALRGLEPADRTYCFDGFSCAAETGVALYSLHDLMPADVAEAMARHGXTLUYA 180
DB 122 CRSSALRGLEPADRTYCFDGFSCAAETGVALYSLHDLMPADVAEAMARHGXTLUYA 181
QY 181 LHPPEVLLPQGYHTTSYLLIHGDRAVYTYGDSAGNHDVSLIRAMIRTKLVGH 240
DB 182 LHPPEVLLPQGYHTTSYLLIHGDRAVYTYGDSAGNHDVSLIRAMIRTKLVGH 241
QY 241 PLVIERVRAIGCHFVLLTLAAPPSPMPVYPRSTEVYVRSIFGSGPSLEPPSACSTK 300
DB 242 PLVIERVRAIGCHFVLLTLAAPPSPMPVYPRSTEVYVRSIFGSGPSLEPPSACSTK 301
QY 301 STPHAVPVIHWRMLFEGATLDDQACSRMTYLGISTYKYTVGLVANEENASEDL 360
DB 302 STPHAVPVIHWRMLFEGATLDDQACSRMTYLGISTYKYTVGLVANEENASEDL 361
QY 361 TAXITAAVLTICHOYLRTQATSKGMRRLGVEHAQKFTRLYSMLPEKSGROYIPGRO 420
DB 362 TAXITAAVLTICHOYLRTQATSKGMRRLGVEHAQKFTRLYSMLPEKSGROYIPGRO 421
QY 421 FTAQCRRLWSAGFHLDPRVLVFEDESVCRCRTFLKKVAKFCCFMRLQOECTCFLEPAE 480
DB 422 FTAQCRRLWSAGFHLDPRVLVFEDESVCRCRTFLKKVAKFCCFMRLQOECTCFLEPAE 481
QY 481 GIVGDCHNENVESEVDPAEPANHDVGTAAVHGHOLEALYRALNVQDLAARSRIT 540
DB 482 GIVGDCHNENVESEVDPAEPANHDVGTAAVHGHOLEALYRALNVQDLAARSRIT 541
QY 541 ATVEELVSPDLRECRFTVLNKTFTTVVDGALHEANGPEQYVLSFEDASQSMGASHSLT 600
DB 542 ATVEELVSPDLRECRFTVLNKTFTTVVDGALHEANGPEQYVLSFEDASQSMGASHSLT 601
QY 601 YELTPAGLOVRISNSGLDCTATPPPGAGSAPAEVAAFCALYRYNRFTQRHSLTGLM 660
DB 602 YELTPAGLOVRISNSGLDCTATPPPGAGSAPAEVAAFCALYRYNRFTQRHSLTGLM 661
QY 661 LHPPEGLIGFPPSPGHIMESANPCGECTLTTRIMS-TSGSSDPS-PEAANA 714
DB 662 LHPPEGLIGFPPSPGHIMESANPCGECTLTTRIMS-TSGSSDPS-PEAANA 721
QY 715 MATPGLPSTPPVSDIWLPPPESEFYQDAVPP-PADPAGLPGVYLTGPPPPVYAK 773
DB 722 RAATPTL-----AAPLPSPAD-----LSTPSPAPADE 750
QY 774 PS-----IPPSNRRLLYTYPPDGAKYVAGSLFESDCDWLVNASNPGHRPGGGLC 823
DB 751 PASGATAGAPALTHOARHRLFTYPPDGSKVFAAGLPESTCTWLVNASNADHRRGGGLC 810
QY 824 HAFYQRPFAFTPTFIMEGLAATTLTPRLIHAVAPRYAEONPKRLAAYRETCSRR 883
DB 811 HAFYQRPFAFTPTFIMEGLAATTLTPRLIHAVAPRYAEONPKRLAAYRETCSRR 870
QY 884 GFAAYPLGSGTYOYVPSLPPDAMERNHRPGDELVTPEAAMPFANKPAQCVLITTEBT 943
DB 871 GFAAYPLGSGTYOYVPSLPPDAMERNHRPGDELVTPEAAMPFANKPAQCVLITTEBT 930
QY 944 ARTANLAEIDATEVGRACAGCTTSPGIVHYOTFAGVGSKRSIOGDDVYVVVPT 1003
DB 931 ARTANLAEIDATEVGRACAGCTTSPGIVHYOTFAGVGSKRSIOGDDVYVVVPT 990
QY 1004 ELRNSWRRRGFAFTPTHTAARVYIDAPSLPHLLILHMQRASVHLGDPNOI 1063
DB 991 ELRNSWRRRGFAFTPTHTAARVYIDAPSLPHLLILHMQRASVHLGDPNOI 1050
QY 1064 PAIDFEHAGLVPAIRPELAPTSWMVYTHRCPADVCELLRGAPKTIQTSRVLRSLFWNP 1123
DB 1051 PAIDFEHAGLVPAIRPELAPTSWMVYTHRCPADVCELLRGAPKTIQTSRVLRSLFWNP 1110
QY 1124 AIGKLVYTOAKKANPGATVHEAGATFTFTTIATADAGLIQSSRAHIVALTTRHT 1183

Db	1111	AVGKLVFTQAARAANGSTVTHAEAGATYTTETTTTATADAGLQSSRAHAIVALTHT	11770
Qy	1184	EKCVILDPAGLLREVGISDVIVNFFLAGEGVXHRPSVIPRGNDONLGTLOAFPSCQ	1243
Db	1171	EKCVIIDAPGLREVGISDVIIVNFFLAGELIGHQRPVYIPRGNDPANDVTLAAPPSCQ	1230
Qy	1244	ISAHQIAELFGRHPAVAAVILPCCLEQGLLYLPQELTDSVTELDIIVCRMMA	1303
Db	1231	ISAHQIAELFGRHPAVAAVILPCCLEQGLLYLPQELTDSVTELDIIVCRMMA	1290
Qy	1304	PSQKAVLSTLVGRYGRRTKLYEAHSDVRESLAREPTIGVQATTCELYELVEAMEK	1363
Db	1291	PSQKAVLSTLVGRYGRRTKLYEAHSDVRESLAREPTIGVQATTCELYELVEAMEK	1350
Qy	1364	GQGSAAVLIEDLCNRDYSRTTFEQKXCNFTTGETIAGHKVQGISAMSKTFCALFGEF	1423
Db	1351	GQGSAAVLIEDLCNRDYSRTTFEQKXCNFTTGETIAGHKVQGISAMSKTFCALFGEF	1410
Qy	1424	RAIKETALLPNIPIFGDYEESVAAVAVSGSGCMVENFSEFSDONNFSGLGCV	1483
Db	1411	RAIKETALLPQGVTFGDYEDTVAASAAPAAKAAWVENFSEFSDONNFSGLGCA	1470
Qy	1484	VMEECGPMWLIRLYHLYRSAMILQAPKESLKGFMKKHSGEPTLLMTVNMALIAHCY	1543
Db	1471	IMECCGPMWLIRLYHLYRSAMILQAPKESLKGFMKKHSGEPTLLMTVNMALIAHCY	1530
Qy	1544	EFRRFRRAAKRGDSVYLCDYDQSRNMAALLIAGCLKLVYRPIGLYAGVYVAPGIGT	1603
Db	1531	DFRQLQAAARKGDSVYLCEYYSQSGAAVLLIAGCLKLVYRPIGLYAGVYVAPGIGA	1590
Qy	1604	LPDVVRFRAGRLSEKMGMPGGERAEQRLAVCFRLGTLVNAOCVADVVSRYVGSPLGVH	1663
Db	1591	LPDVVRFRAGRLTEKMGMPGERAEQRLAVSDFRLKTLVNAQCMADVVSRYVGSPLGVH	1650
Qy	1664	NLIGMLDTADGKAHFTETIKPVLDTNSTIQAYE 1698	
Db	1651	NLIGMLQAVADGKAHFTESYKPVLDLTNSTICRYE 1685	
RESULT 6			
Q0	096410		
AC	069410	PRELIMINARY;	PRT; 1693 AA.
DT	01-NOV-1996 (TEMBLrel. 01, Created)		
DT	01-JAN-1998 (TEMBLrel. 05, last sequence update)		
DT	01-JUN-2001 (TEMBLrel. 17, last annotation update)		
DE	METHYL TRANSFERASE.		
OS	Hepatitis E virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage.		
OX	NCBI TaxID=12461;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=HEV037;		
RA	Donati M.C., Fagan E.A., Harrison T.J.;		
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; X98292; CA66936.1; -		
DR	InterPro: IPR000606; Viral_helicase1.		
DR	InterPro: IPR002588; V_methyltransf.		
DR	InterPro: IPR002589; DUF27.		
DR	Pfam: PF01443; Viral_helicase1; 1.		
DR	Pfam: PF01660; Vmethyltransf; 1.		
DR	Pfam: PF01661; DUF27; 1.		
DR	SMART: SM00506; Alpp; 1.		
Q0	SEQUENCE 1693 AA; 185190 MW; C347C2A36DEDB9B23 CRC64;		

Query Match	82.9%;	Score 7468;	DB 12;	Length 1693;
Best Local Similarity	82.3%;	Pred. No. 0;		
Matches 1402;	Conservative 109;	Mismatches 167;	Indels 26;	Gaps 7;
OY	1	PGTTATFEQALAAANSALANAVVVRPFPSKQTETLLINLMPRQTVRPEVIAWHPIOR	60	

D	b	10	P	G	I	T	A	E	O	A	L	A	N	S	A	N	A	N	V	V	P	E	L	S	H	O	I	E	L	L	M	O	P	R	O	L	V	R	E	P	E	V	M	N	H	P	O	R		69								
Q	y	61	V	I	N	H	E	L	O	Y	C	R	A	R	A	G	R	C	L	E	V	A	H	P	R	S	I	N	P	N	V	L	H	C	F	L	R	P	G	R	D	V	O	R	N	S	A	P	T	R	G	P	A	N	120			
D	b	70	V	I	N	H	E	L	O	Y	C	R	A	R	A	G	R	S	C	R	C	L	E	I	G	A	H	P	R	S	I	N	P	N	V	H	C	F	L	R	P	G	R	D	V	O	R	N	T	A	P	T	R	G	P	A	N	129
Q	y	121	C	R	S	A	L	G	L	P	P	A	D	R	T	C	F	D	F	S	C	A	F	A	E	T	G	A	L	S	L	H	D	L	M	P	A	D	V	A	E	A	M	A	R	I	G	T	R	L	Y	A	180					
D	b	130	C	R	S	A	L	G	L	S	A	A	D	R	T	C	F	D	F	S	C	N	P	A	E	T	G	A	L	S	L	H	D	M	S	P	D	V	A	E	A	M	F	R	G	M	T	R	L	Y	A	189						
Q	y	181	L	H	I	P	E	L	L	P	G	R	N	T	S	T	S	T	L	H	D	G	R	A	V	T	E	G	D	S	A	C	N	D	S	L	I	A	M	I	R	T	T	K	I	V	S	D	H	240								
D	b	190	L	H	I	P	E	L	L	P	G	R	N	T	S	T	S	T	L	H	D	G	R	A	V	T	E	G	D	S	A	C	N	D	S	L	I	A	M	I	R	T	T	K	I	V	S	D	H	249								
Q	y	241	P	L	T	E	R	R	A	I	G	C	H	F	V	L	L	T	A	P	E	S	P	M	P	V	P	R	S	T	E	V	N	S	I	F	P	G	S	P	L	E	P	S	A	C	T	K	300									
D	b	250	P	L	T	E	R	R	A	I	G	C	H	F	V	L	L	T	A	P	E	S	P	M	P	V	P	R	S	T	E	V	N	S	I	F	P	G	S	P	L	E	P	S	A	C	T	K	309									
Q	y	301	S	T	F	H	A	V	E	H	I	M	D	R	L	M	F	G	A	T	L	D	O	A	E	C	C	S	R	L	M	T	Y	L	G	I	S	K	Y	V	T	A	L	A	N	E	G	M	N	S	E	D	L	360				
D	b	310	S	T	F	H	A	V	E	H	I	M	D	R	L	M	F	G	A	T	L	D	O	A	E	C	C	S	R	L	M	T	Y	L	G	I	S	K	Y	V	T	A	L	A	N	E	G	M	N	S	E	D	L	369				
Q	y	361	T	A	X	T	A	A	L	T	I	C	H	O	R	I	T	R	O	A	I	S	K	M	R	L	C	V	E	H	A	O	K	F	T	R	L	X	S	W	L	E	P	S	G	R	I	T	G	R	O	L	420					
D	b	370	T	A	X	T	A	A	L	T	I	C	H	O	R	I	T	R	O	A	I	S	K	M	R	L	C	V	E	H	A	O	K	F	T	R	L	X	S	W	L	E	P	S	G	R	I	T	G	R	O	L	429					
Q	y	421	F	Y	A	C	R	K	M	L	S	G	F	L	D	R	V	L	P	E	D	S	P	C	R	T	P	L	K	V	A	C	C	F	M	R	L	G	O	E	C	T	C	L	E	A	E	480										
D	b	430	F	Y	A	C	R	K	M	L	S	G	F	L	D	R	V	L	P	E	D	S	P	C	R	T	P	L	K	V	A	C	C	F	M	R	L	G	O	E	C	T	C	L														

```
Db 1070 PAIRPDLAFTSMWHVTHRCPADVCELLRGAVPMIOITTSRVLSLNGEPEAVGOKIVETQA 1129
Oy 1135 AKAANGATVTHEAGATFETTTIATADARGLIOSSRAHAVALRHTKECVILDPAGL 1194
Db 1130 AKAANGSVTHEAGATFETTTIATADARGLIOSSRAHAVALRHTKECVIIDAPGL 1189
Oy 1195 LKEVGISDIYVNNFVLAGEVGHKRPVYPRGNPDONLGLQAFPPSCISAVHQLAEEL 1254
Db 1190 LKEVGISDAIVNNFVLAGEIGIHPORSVIPRGPNPTNVDTLAAPPSCOISAFHOIAEEL 1249
Oy 1255 GHRPAPVAANLPPCPLEGLLMLYMPDELVSVDVLEFLDVIHCHMAAPSQOKAVLSL 1314
Db 1250 GHRPAPVAANLPPCPLEGLLMLYMPDELVSVDVLEFLDVIHCHMAAPSQOKAVLSL 1309
Oy 1315 VGRYGRRTLYEAHSDVSESLARFPIPIGYOATTCLELYEAVAEKGGDSAVLELD 1374
Db 1310 VGRYGRRTLYNAHSDVSDVSLARFPIPIGYOATTCLELYEAVAEKGGDSAVLELD 1369
Oy 1375 LCNRDVSRITFFOKCKNKTGTGTTAHHGVGOGISAMSTFCALFGPMFRAIEKETALL 1434
Db 1370 LCNRDVSRITFFOKCKNKTGTGTTAHHGVGOGISAMSTFCALFGPMFRAIEKETALL 1429
Oy 1435 PNIFFGDAEESVFAAASGASGVFENDSEPDSTONNLSIGECVYMECGMPQL 1494
Db 1430 PGGVITGDADFDDIVEAATAAKASVAFENDSEPDSTONNLSIGECVYMECGMPQL 1489
Oy 1495 IRLYHLVRSAMTLOAPKESLKGFWKHSGEPTLLMNTVWNNALIAHCYEFDRVAAR 1554
Db 1490 IRLYHLVRSAMTLOAPKESLKGFWKHSGEPTLLMNTVWNNALIAHCYEFDRVAAR 1549
Oy 1555 GDSVYVLCSDIYQSRNAALINGCGILKLVDRPIGLYAGVVAAPGLTLPDVVPRAGL 1614
Db 1550 GDSVYVLCSEYKQSGAAALINGCGILKLVDRPIGLYAGVVAAPGLTLPDVVPRAGL 1609
Oy 1615 SEKNNGPGERAEOLRLAVCDLRLGTNTVAQVCDPVVSVVSPGLVNNLGMQTIAD 1674
Db 1610 SEKNNGPGERAEOLRLAVCDLRLGTNTVAQVCDPVVSVVSPGLVNNLGMQTIAD 1669
Oy 1675 GKAFHETIKPVLDTLNSIIQRIE 1698
Db 1670 GKAFHETIKPVLDTLNSIIQRIE 1693

RESULT 7
O9MIL5 PRELIMINARY: PRT: 1693 AA.
AC O9MIL5:
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN.
OC Hepatitis E virus.
OS Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK15/92;
RX MEDLINE=99049628; PubMed=9833882;
RA Gouvea V.; Snellings N., Popok M.J., Longer C.F., Innis B.L.;
RT "Hepatitis E virus: complete genome sequence and phylogenetic analysis
of a Nepali isolate."
RL Virus Res. 57:21-26(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TK15/92;
RA Gouvea V.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF051830; AAC97186.1;
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1.
DR Pfam: PF01660; Vmethyltransf.1.
```

```
DR Pfam: PF01661; DUF27.1.
DR SMART: SM00506; A1pp; 1.
KW Polyprotein.
SQ SEQUENCE 1693 AA; 185349 MW; A895ACD1DAE2PBD CRC64;

Query Match 82.7%; Score 7457; DB 12; Length 1693;
Best Local Similarity 81.7%; Pred. No. 0;
Matches 1397; Conservative 110; Mismatches 166; Indels 36; Gaps 6;

Oy 1 PGTTTAIEQAALAAANSALANAVVPRPSRQTELLINLMQPROLYEFREVLMMHPIOR 60
Db 10 PGTTTAIEQAALAAANSALANAVVPRPSRQTELLINLMQPROLYEFREVLMMHPIOR 69
Oy 61 VHNNELEQCRARAGRCLEVGAPRSINDPNVLRFCFLPRGRDQORYVSAPTRGPAAN 120
Db 70 VHNNELELYCRARSGCLEIGAHPRISINDPNVLRHCFILPRGRDQORYVTAPTRGPAAN 129
Oy 121 CRRSALRGLPPADRTYCFDGFSCAFPAETGALYSLHDLMPADVAEMARRHGXRLYAA 180
Db 130 CRRSALRGLPPADRTYCFDGFSCAFPAETGALYSLHDLMPADVAEMARRHGXRLYAA 189
Oy 181 LHLPEVLLPCTYHTSTYLLIHDGRVAVYEGTSAQYNHDVSLIAMIWITKIVGH 240
Db 190 LHLPEVLLPCTYHTSTYLLIHDGRVAVYEGTSAQYNHDVSLIAMIWITKIVGH 249
Oy 241 PLVIERVRAIGCHFVLLTFAAPESPMPVYPRSTREYVRSIFGSGSPSLFSPASCSTK 300
Db 250 PLVIERVRAIGCHFVLLTFAAPESPMPVYPRSTREYVRSIFGSGSPSLFSPASCSTK 309
Oy 301 STEHAVVPIWDRMLFGATLDDQAFCSSRLMTYLRGISYKYTVGALVANEGMNSDAL 360
Db 310 STEHAVVPIWDRMLFGATLDDQAFCSSRLMTYLRGISYKYTVGALVANEGMNSDAL 369
Oy 361 TXXITTAAYITFCHQYRLTRQALSKGRRLGVEHAQKFTIRLXSMLEKSGRYITGRQO 420
Db 370 TXXITTAAYITFCHQYRLTRQALSKGRRLGVEHAQKFTIRLXSMLEKSGRYITGRQO 429
Oy 421 FVAOCRRMLSAGFHLDPRLVLFDESVCRCRFLKKVAKFCFCFMRMLGQECTCFLEPAE 480
Db 430 FVAOCRRMLSAGFHLDPRLVLFDESVCRCRFLKKVAKFCFCFMRMLGQECTCFLEPAE 489
Oy 481 GLVGDHNDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPDIIAARSLT 540
Db 490 GLVGDHNDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPDIIAARSLT 549
Oy 541 ATVELVASPDLECRFTVGNKTFRTTYVDGALHEANGPQOYVLSDDSQSGAGSHST 600
Db 550 ATVELVASPDLECRFTVGNKTFRTTYVDGALHEANGPQOYVLSDDSQSGAGSHST 609
Oy 601 YELTPAGLOVRISNGLDCTATFPPGGAAPSAPRGVAAFCSCALYVRNRTORHSLTGLGM 660
Db 610 YELTPAGLOVRISNGLDCTATFPPGGAAPSAPRGVAAFCSCALYVRNRTORHSLTGLGM 669
Oy 661 LHPBGLIGTFPPSPGHIMESANPFCGEGTLYTRTWS-TSGFSDSPS-----PRAAPA 714
Db 670 LHPBGLIGTFPPSPGHIMESANPFCGEGTLYTRTWS-TSGFSDSPS-----PRAAPA 729
Oy 715 MAATGCLPHSTIPPVSDIWLPPPSSEFOVDAAPV-PPAPDPAGLGPPVYLTPPPPPVHK 773
Db 720 MAATGCLPHSTIPPVSDIWLPPPSSEFOVDAAPV-PPAPDPAGLGPPVYLTPPPPPVHK 782
Oy 774 PSIP-----PPSRNRLLTYPDGAIVYAGSLFESDCMLVYVNSNGHPPGGGICHAFAOR 829
Db 765 PSIP-----PPSRNRLLTYPDGAIVYAGSLFESDCMLVYVNSNGHPPGGGICHAFAOR 838
Oy 830 PPEAFYPTFEIIRREGIAAYTLTPRIIAHVAAPDVEEONPKRLAAAYRETCSRGGTAAP 889
Db 825 PPEAFYPTFEIIRREGIAAYTLTPRIIAHVAAPDVEEONPKRLAAAYRETCSRGGTAAP 898
Oy 890 LLGSGIYQVPSLSFADAWERNHRPDELTLTEPAANWEFANKPAOPVLTITEDTARTANL 949
Db 885 LLGSGIYQVPSLSFADAWERNHRPDELTLTEPAANWEFANKPAOPVLTITEDTARTANL 958
```

```

OY 950 ALEIDATATEGRACACCTISPGIVHIOFTAGVPGSKSRISIOGDVYVYVPTRELNSM 1009
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 945 ALEIDATIDVGRACACGCRTPGVYOQFTAGVPGSKSRISIOADVVYVPTRELNSM 1004
OY 1010 RRGFAFPHTARVTRGRVYIDEAPSLPHLLLLHMQRASSVHLGDPPNIPIDPE 1069
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1005 RRGFAFPHTARVTRGRVYIDEAPSLPHLLLLHMQRAVHLGLDPPNIPIDPE 1064
OY 1070 HAGLVAIRPELAPTSMMVYTHRCPADVCELKAGAPKIQTYSRVLSLFWNPATGOKL 1129
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1065 HAGLVAIRPELAPTSMMVYTHRCPADVCELKAGAPKIQTYSRVLSLFWNPATGOKL 1124
OY 1130 VYTOAKAANPGATVYHEAGATFTETTTIATDANGLIOSRAHAIVALTRTEKCVIL 1189
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1125 VFTQAKKANPGSVTVHEAGATFTETTTIATDANGLIOSRAHAIVALTRTEKCVIL 1184
OY 1190 DAGGLREAGISDVIYNNEFLAGEGVXHRPSVTPRGNPDPNLTGLOAPPSCOISAYHQ 1249
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1185 DAGGLREAGISDVIYNNEFLAGEGVXHRPSVTPRGNPDPNLTGLOAPPSCOISAYHQ 1244
OY 1250 LAELGHRAPVAAVLPCEPELEOGLLYMPQELVSDSVLVEFELTDIVHCRMAPSQRKA 1309
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1245 LAELGHRAPVAAVLPCEPELEOGLLYMPQELVSDSVLVEFELTDIVHCRMAPSQRKA 1304
OY 1310 VLTIVGRGRRTKLYEAHSDVRESLARIPTIGVQATTCGLYELVEMAEKGDGSA 1369
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1305 VLTIVGRGRRTKLYEAHSDVRESLARIPTIGVQATTCGLYELVEMAEKGDGSA 1364
OY 1370 VLELDLCNDRVSRITFEOKCKNKFTTGETTAHGVKGGISAMSKTCALFGMPFRAIEKE 1429
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1365 VLELDLCNDRVSRITFEOKCKNKFTTGETTAHGVKGGISAMSKTCALFGMPFRAIEKE 1424
OY 1430 IIALLPNIFGDAYEESVFAAASGAGSCMVEENFSEEDSTONNESLIGECVMEECG 1489
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1425 IIALLPNIFGDAYEESVFAAASGAGSCMVEENFSEEDSTONNESLIGECVMEECG 1484
OY 1490 MPQWLLRHLVLSAMTLOAPKESLKGFMKHSGEPTLLMNTVMNAALIAHCYEFEDFR 1549
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1485 MPQWLLRHLVLSAMTLOAPKESLKGFMKHSGEPTLLMNTVMNAALIAHCYEFEDFR 1544
OY 1550 VAAFKGDDSVLCSYDROSNAALATAGCLKLVDRPIGLIYAGVVAAGLGLTLPDVR 1609
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1545 VAAFKGDDSVLCSYDROSNAALATAGCLKLVDRPIGLIYAGVVAAGLGLTLPDVR 1604
OY 1610 FAGRLSEKMGPGPERAEQLRLAVCDFLRLTNVAQVCDVVSRYGVSPGLVHNLIGML 1669
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1605 FAGRLSEKMGPGPERAEQLRLAVCDFLRLTNVAQVCDVVSRYGVSPGLVHNLIGML 1664
OY 1670 QTADGKAHFTETIKPYLDLTNSIIORVE 1698
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1665 QAVADGKAHFTESVKPYLDLTNSIWCVE 1693

```

```

RESULT 8
O89444 PRELIMINARY: PRT: 1693 AA.
AC O89444;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE UNNAMED PROTEIN PRODUCT.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RA vin S.R., Purcell R.H., Emerson S.U.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K52-87;
RX MEDLINE=95176571; PubMed=7871758;

```

```

RA Yin S., Purcell R.H., Emerson S.U.;
RT "A new Chinese isolate of hepatitis E virus: comparison with strains
RT recovered from different geographical regions.";
RL Virus Genes 9:23-32(1994).
DR EMBL: L25547; AAA91078.1; -.
DR EMBL: L25595; AAA65488.1; -.
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; Viral_helicase1.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1.
DR Pfam: PF01660; Vmethyltransf. 1.
DR Pfam: PF01661; DUF27; 1.
DR SMART: SM00506; Alpp; 1.
SQ SEQUENCE 1693 AA; 185122 MW; 53914B8302338D5E CRC64;

```

```

Query Match 82.7%; Score 7456; DB 12; Length 1693;
Best local similarity 81.8%; Pred. No. 0;
Matches 1401; Conservative 108; Mismatches 161; Indels 42; Gaps 8;

```

```

OY 1 PGTTAIEQALAAANSALANAVVPELSRVOTEIILINIMOPROLYFREVLMNPIOR 60
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 10 PGTTAIEQALAAANSALANAVVPELSRVOTEIILINIMOPROLYFREVLMNPIOR 69
OY 61 VINHELEOYCARAGRCLEVGAPRSTINDPNVLRHCFELRPGVGRDORWT SAPTPRPAAN 120
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 70 VINHELEOYCARAGRCLEVGAPRSTINDPNVLRHCFELRPGVGRDORWT SAPTPRPAAN 129
OY 121 CRRSALRGLPPADRTYCFDFSGCAFAETGVALYSLHMDLPADVABAMA RHGXTRLYAA 180
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 130 CRRSALRGLPPADRTYCFDFSGCAFAETGVALYSLHMDSPDVABAMRHGXTRLYAA 189
OY 181 LHLPEVLEPPGTYHTTSYLLHDGDAVYTTESDTSAGNHDVSLIRAITRTTKYGDH 240
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 190 LHLPEVLEPPGTYHTTSYLLHDGDAVYTTESDTSAGNHDVSLIRAITRTTKYGDH 249
OY 241 PLVIERVRAIGCFVLLTAPRSPMPVYPYPRSTEVYRSIRPGSGSLSPFSACSTK 300
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 250 PLVIERVRAIGCFVLLTAPRSPMPVYPYPRSTEVYRSIRPGSGSLSPFSACSTK 309
OY 301 STEFAVPHIWMRLMFGATLDDOAFCCSRIMTYLRGISYKVYGVALVANEWNASBDAL 360
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 310 STEFAVPHIWMRLMFGATLDDOAFCCSRIMTYLRGISYKVYGVALVANEWNASBDAL 369
OY 361 TAYTTAAVLTICHOVRTIRTAISKGMRLGVENAKITRLYSMLFEKSGRDYIPGQOL 420
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 370 TAYTTAAVLTICHOVRTIRTAISKGMRLGVENAKITRLYSMLFEKSGRDYIPGQOL 429
OY 421 FYAQCRWLSAGFLDPRVLVFDSEVPCRCRTFLKTVAGKFCFCFMRMLGOECCTFLPAE 480
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 430 FYAQCRWLSAGFLDPRVLVFDSEVPCRCRTFLKTVAGKFCFCFMRMLGOECCTFLPAE 489
OY 481 GLVGDHNDENAYEGSEVDEPAEPAHLDVSGTYAVHGOELALYBALNVPDIIAASRLT 540
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 490 GLVGDHNDENAYEGSEVDEPAEPAHLDVSGTYAVHGOELALYBALNVPDIIAASRLT 549
OY 541 ATVELVASPDLRECTVYLGKNTFRITYDGAHLEANPEOYVLSFDSRSGMGSGHSIT 600
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 550 ATVELVASPDLRECTVYLGKNTFRITYDGAHLEANPEOYVLSFDSRSGMGSGHSIT 609
OY 601 YELTPAGLOYRISNGDCTATPPGGAAPSAPAEVAAFCALRYRFRTHSLTJGGLW 660
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 610 YELTPAGLOYRISNGDCTATPPGGAAPSAPAEVAAFCALRYRFRTHSLTJGGLW 669
OY 661 LHPGLGIPPPFSPGHIMESANPFCGEGLYLTRWS-----TSGFSSDFSPPE 709
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 670 FHPGGLGIPFAPFSPGHIMESANPFCGEGLYLTRWSVDAVSSPAQDGLGFTSEPSIPS 729
OY 710 AAPAMAAATPGLPHSTPVSVDIYVLPPEPSEFQVDAAPVPPAP---DPAGLPGEVYVLTLP 766
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 730 RAA---TLTPAAP-----LPPAP---DPSPTESAPARGEP---PG---ATAG 766
OY 767 PPPVHNPSTIPPPSRNRRLTYTPDGAKYVAGSLFESDCDMLVANSNPGHRGGGLCHAF 826

```

Db 767 AAPTITQ-----TARHRLFTYPDOSKVFAGSLFESTCWLWNASVNDHRPGEGELCHAF 821
QY 827 YORFPAFYPTFEIMEGELAAYLTLPRIITHAAPDYRVBONKRLAEAYREFCSRGTA 866
Db 822 YORFPAFSDAASVMDGAAATLTLPRIITHAAPDYRLHNKRLAEAYREFCSRGTA 861
QY 887 AYPPLGSGIYQVPSLSPDAMERNHRPGDELXYLTPBAWMEANKPAQVLTITTEDTART 946
Db 882 AYPPLGSGIYQVPSLSPDAMERNHRPGDELXYLTPBAWMEANKPAQVLTITTEDTART 941
QY 947 ANLALEIDATVEGRACAGCTISPGIVHQFTAGVSGSKSIOGDDVYVVPFREL 1006
Db 942 ANLALEIDATVEGRACAGCTISPGIVHQFTAGVSGSKSIOGDDVYVVPFREL 1001
QY 1007 NSWRRGFAFTPHTAARTIGRRVIDEAPSLPHLLLLHMO RASSVHLGDPNCPAI 1066
Db 1002 NSWRRGFAFTPHTAARTIGRRVIDEAPSLPHLLLLHMO RASSVHLGDPNCPAI 1061
QY 1067 DEHAGLVPAIRPELAPTSWXYVTHRCPADVCELIRGAYPKIOTTSRVLSLFWNEPATG 1126
Db 1062 DEHAGLVPAIRPELAPTSWXYVTHRCPADVCELIRGAYPKIOTTSRVLSLFWNEPATG 1121
QY 1127 OKLVYQAANKANPGAIYVHEAGATFETTTIATADAGLIOSRAHAIVALTHTKXC 1186
Db 1122 OKLVYQAANKANPGAIYVHEAGATFETTTIATADAGLIOSRAHAIVALTHTKXC 1181
QY 1187 VILDAFGLREVGISDVIVNNFPLAGGEVGHXHSYIIPGNPNONIGTLOAPPSOISA 1246
Db 1182 VILDAFGLREVGISDVIVNNFPLAGGEVGHXHSYIIPGNPNONIGTLOAPPSOISA 1241
QY 1247 YHQLAEELGHRAPVAVALPCPELEBGLLYMPOELVSDVLYFELTJDIVHCRMAPSQ 1306
Db 1242 YHQLAEELGHRAPVAVALPCPELEBGLLYMPOELVSDVLYFELTJDIVHCRMAPSQ 1301
QY 1307 KRAVLSTLVGRGRKTKLYEAHSDVRESLAPITPIGVQATTCCLYELVEAMVEKGD 1366
Db 1302 KRAVLSTLVGRGRKTKLYEAHSDVRESLAPITPIGVQATTCCLYELVEAMVEKGD 1361
QY 1367 GSAVLELDLCNDRVSRTFEOKCKNFTTGETTAHKGVGOGISAMSTFCALGPMFRAI 1426
Db 1362 GSAVLELDLCNDRVSRTFEOKCKNFTTGETTAHKGVGOGISAMSTFCALGPMFRAI 1421
QY 1427 EKEIALLPRTNYGDAYEESVFAAASVAGSCWMEFENFSEPDSTONNESLGLCECVME 1486
Db 1422 EKEIALLPRTNYGDAYEESVFAAASVAGSCWMEFENFSEPDSTONNESLGLCECVME 1481
QY 1487 ECGMPQMLRLYLHVSANITLQAPKESLGFPMKHSGBEGTLLMNTVMNAVITTHCYDPR 1546
Db 1482 ECGMPQMLRLYLHVSANITLQAPKESLGFPMKHSGBEGTLLMNTVMNAVITTHCYDPR 1541
QY 1547 DFRVAAFKGDSDVILCSDRQSRNAALJAGCGLKLVDPRTGLAGVVAAGLTLDP 1606
Db 1542 DFRVAAFKGDSDVILCSDRQSRNAALJAGCGLKLVDPRTGLAGVVAAGLTLDP 1601
QY 1607 VFRFAGRLSEKMMWGPBPBAEOLRLAVCDFLGLINVAOCVVDVSVRYGVSPLVHNL 1666
Db 1602 VFRFAGRLSEKMMWGPBPBAEOLRLAVCDFLGLINVAOCVVDVSVRYGVSPLVHNL 1661
QY 1667 GMLQTIADGKAHTEIKRPVLTLSNIORVE 1698
Db 1662 GMLQTIADGKAHTEIKRPVLTLSNIORVE 1693

RESULT 9
ID 081876 PRELIMINARY: PRT: 1693 AA.
AC 081876:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE COMPLETE GENOME SEQUENCE.
OS Hepatitis E virus.

OC Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_Taxid=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UIC8179;
RA Uchida T.;
RL Submitted (May-1992) to the EMBL/Genbank/DBJ databases.
DR EMBL: D11093; BAA01865.1; -
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1.
DR Pfam: PF01660; Vmethyltransf. 1.
DR Pfam: PF01661; DUF27. 1.
DR SMART: SM00506; A1pp. 1.
SQ SEQUENCE 1693 AA; 184839 MW; 4E422ED55AE0B1 CRC64;

Query Match 82.4%; Score 7424; DB 12; Length 1693;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1397; Conservative 108; Mismatches 165; Indels 42; Gaps 8;

QY 1 PGTTTAIEQAAIAAANSAANAVVVPFLSRVOTELLINLMQPROLVFPRVLMNPIQR 60
Db 10 PGTTTAIEQAAIAAANSAANAVVVPFLSHOOIELINLMQPROLVFPRVLMNPIQR 69
QY 61 VTHNELEOYCRARAGRCLEVGAPRSINDPNVYLHRCFLRPVGRDQVRYVSAPTGPAN 120
Db 70 VTHNELELYCRARAGRCLEIGHAPRSINDPNVYVHRCFLRPAGRDQVRYVSAPTGPAN 129
QY 121 CRRSALRGLPADRTYCFDGFSCAFPAETGVALYSLHDLMPDAVAAARHGXTLYAA 180
Db 130 CRRSALRGLPADRTYCFDGFSCAFPAETGVALYSLHDLMPDAVAAARHGXTLYAA 189
QY 181 LHPREVLLPRTYHTSTLLIHDGRVAVYTBGTSAGYNDVSLRLMIRTKIVGCH 240
Db 190 LHPREVLLPRTYHTSTLLIHDGRVAVYTBGTSAGYNDVSLRLMIRTKIVGCH 249
QY 241 PLYIEVRAIGCFHVLTLAAPEPSPMPVYPRSTEVYVRSIFGGSPLSPFSACTRK 300
Db 250 PLYIEVRAIGCFHVLTLAAPEPSPMPVYPRSTEVYVRSIFGGSPLSPFSACTRK 309
QY 301 STEHAVVHIMRLMFGATLDDQAFCCSRMLTYLIGISYKVTGVALVANEWNASEDAL 360
Db 310 STEHAVVHIMRLMFGATLDDQAFCCSRMLTYLIGISYKVTGVALVANEWNASEDAL 369
QY 361 TAXITAAVLTICHORLRTQALSKGRRLGVEHAQFTRILYSMLFEKSGROYIPROQ 420
Db 370 TAXITAAVLTICHORLRTQALSKGRRLGVEHAQFTRILYSMLFEKSGROYIPROQ 429
QY 421 FYAQRRLWSAGFHLDPRLVLFDESVPQCRRTFLKVKAGFCCFMRLGOECTCELEPAE 480
Db 430 FYAQRRLWSAGFHLDPRLVLFDESVPQCRRTFLKVKAGFCCFMRLGOECTCELEPAE 489
QY 481 GLVGDHNDNEAYEGSEVDPAEPALDVSCTAIVHSHOLEALYRALNVPDIAAARSRLT 540
Db 490 GLVGDHNDNEAYEGSEVDPAEPALDVSCTAIVHSHOLEALYRALNVPDIAAARSRLT 549
QY 541 ATVELVAASPDRLCORTVLTGNTKTRTYVNDGAHLEANGPPOYLSPFASROSQAGHSJT 600
Db 550 ATVELVAASPDRLCORTVLTGNTKTRTYVNDGAHLEANGPPOYLSPFASROSQAGHSJT 609
QY 601 YELTPAGLOVRISSNGLDCTATFPFGAGAPAAAGVAAACSAALYRYNRTQHSLSLGLAM 660
Db 610 YELTPAGLOVRISSNGLDCTATFPFGAGAPAAAGVAAACSAALYRYNRTQHSLSLGLAM 669
QY 661 LHPREGLLTGTPPSPGHITWESANPCGEGTLYTRWS-----TSGSSDSFSPPE 709
Db 670 LHPREGLLTGTPPSPGHITWESANPCGEGTLYTRWS-----TSGSSDSFSPPE 729
QY 729 LHPREGLLTGTPPSPGHITWESANPCGEGTLYTRWS-----TSGSSDSFSPPE 766
Db 730 LHPREGLLTGTPPSPGHITWESANPCGEGTLYTRWS-----TSGSSDSFSPPE 766

QY	121	CRBSALGLPPADDTTCGDFGSCAFAEYGVALYSLHDLMPADVAEAMARHGXTRLTA	180
Db	130	CRSRLGLPPADDTTCGDFGSCNPPAEFGVALYSLHDMSPDVAEAMFRRGMRLTYAA	189
QY	181	LHLPEVLLPFGYVHTSYLLIHDGQRAVVTYEGDTSAGYNHDVSLTRAMIRTKLVGDH	240
Db	190	LHLPEVLLPFGYRTASLYLLIHDGRVAVVTYEGDTSAGYNHDVSLNLRSMITTKVYGDH	249
QY	241	PLVIERVAIGCHEVLLTTAAPESPMPVVPVPRSTEVYVRSIFGPGSPSLFPSCSTK	300
Db	250	PLVIERVRAIGCHEVLLTTAAPESPMPVVPVPRSTEVYVRSIFGPGSPSLFPSCSTK	309
QY	301	STFHAIVVHTMDRLMLFGALDDQACCSMLMTYLGISYKYTVGALVANBGMWNSBDAL	360
Db	310	STFHAIVVAHITMDRLMLFGALDDQACCSMLMTYLGISYKYTVGTLVANBGMWNSBDAL	369
QY	361	TAXITAAVLTICHOVRTYRTQAIISKMRRLGVEHAQEFIRLWSLPEKSGROYIGROLO	420
Db	370	TAVITAAVLTICHOVRTYRTQAIISKMRRLEREHAQEFITRLWSLPEKSGROYIGROLO	429
QY	421	FYAACRMLWSAGFLHDPRLVLPFDESVPBCRRTFLKLVAGKFFCCFMRLGQECTCELEPAE	480
Db	430	FYAACRMLWSAGFLHDPRLVLPFDESAPCHCRTAIRAVSKFCOPMKMLGQECTCELOPAE	489
QY	481	GLVDDHGDHNDAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYKALYNQDIAAASRLT	540
Db	490	GAVDDHGDHNDAYEGSDVDAEESALDISISYVVPCTALOPLQALDPAEIVAAASRLT	549
QY	541	ATVEIVASPDLEECRTYLAKGKTRFTYVVDANILEANGPPOVYLSPFASQOSMGASHSLT	600
Db	550	ATVAVSQVDGRIDCELTILGKKTRTSYVDQAVLETNGPERHNLSEFASQSTYAAAPFSILT	609
QY	601	YELTPAGLOVRISSNGDCTATCTPPPGGAPSAPEGEVAAECALYRYNREPTORHSLTGGLW	660
Db	610	YAAASAGLEVRVYVAGGJHDHRAVPAPGVPSPASAGEVTACSALYRPNRAQSLSLTGNMW	669
QY	661	LHPGGLLGITPPSPSGHIMWSANPFCGEGELTYRTKMS-TSGSSDSFP-----PRAAPA	714
Db	670	FHPGGLLGTPPAPSPSGHIMWSANPFCGEGESTLYRTKMS-EVDAVSSPAQDPLGFTISBPIS	729
QY	715	MAAPRGLPHSTPPVSDIWLVPPESEFQVDAAPV-PPADDPGLP-GRVLTLPPEPPVH	772
Db	730	RAAT-----LTP-----AAPLPPAADPSPTPSARAGSAPAGATAR	766
QY	773	KPSIP-PPSRNRLLYTYPDGAIVAGSLFESDCDMLVNASNPGRHGGGLCHAFYQORP	831
Db	767	APATTHQAAHRRRLFTYPDGSKVAFGSLFESTCTMLVNASVNDHRPGGGLCHAFYQORP	826
QY	832	EAFPTPEFIRBEGLAATLTLPRIIHAIVADRYVEDONPKRLAARETCSRGCTATYPL	891
Db	827	ASFPMAASFVRKDGAAATLTLPRIIHAIVADRYLEHNPMLEAAARETCSRGLTAAYPL	886
QY	892	GSGIYOVVSLSPFAMRNHRPGDELTYLPEPAAWMEAKPQOVLTITEDTAPRANAL	951
Db	887	GTCGIYQVPGESFDMERKNRPGDELTYLELAKWMEAKRNPCTLTITTEDAKRANAL	946
QY	952	EIDATAEVGRACAGCTSPGIVAYQCTAGVPGSGSKRSIQGDVVVVVPTRELTNSMR	1011
Db	947	ELDSATDVGRACAGCARTPEVVOYQCTAGVPGSGSKRSITQGDVVVVVPTRELTNSMR	1006
QY	1012	RGEPAATPHTPAARVTTIGRRVVIDEADSLPRHLLLHMORASSVHLGDPNOIPADFEHA	1071
Db	1007	RGEPAAPPHTPAARVTOGRVVIDEADSLPRHLLLHMORAAVHLLGDPNOIPADFEHA	1066
QY	1072	GLVAPATPELAPRSMWVYHRCADVCELRKAYPKIQTSYVLSLSPWNEPAIQOKLVX	1131
Db	1067	GLVAPATPELAPRSMWVYHRCADVCELRKAYPMIQTSYVLSLSPWNEPAVQOKLVF	1126
QY	1132	TOAAKANPAIIVYHEOGATFETETIITATADARGLIOSSRAHAIVALTIRHTEKCVLIDA	1191
Db	1127	TOAAKANPAISVIVYHEOGATYETETIITATADARGLIOSSRAHAIVALTIRHTEKCVLIDA	1186
QY	1192	PGLLRREGISDVIVNNFELAGSEVGHKRSVPIRGNDONLGTLOAFPPSCOISAYHOLA	1251

Query	Match	81.4%: Score 7333; DB 12; Length 1693;	6
QY	1	PEITTAIEGAAALAAANSALANAVVVRPEFSRQTELLINLMQPROLVFPEVLMNHPOR 60	
DB	10	PEITTAIEGAAALAAANSALANAVVVRPEFSRQTELLINLMQPROLVFPEVLMNHPOR 69	
QY	61	VIHNELEOYCRARAGCTLEVGAHPSINDPNVLRHCFRLPVGRDQVRWYSAPTRGPAAN 120	
DB	1187	PGILREVGISDAIVNNEFLAAGGETIGHQRSVYIPKGNPDANVDTLAAPPSCQISAFHQIA 1246	
QY	1252	EEELGNAPVAVAANLPPCPPELEOGELTMPLELTVSDSVLVEFELTDIVHCMAAPSORKAVL 1311	
DB	1247	EEELGNAPVAVAANLPPCPPELEOGELTMPLELTVSDSVLVEFELTDIVHCMAAPSORKAVL 1306	
QY	1312	STLVGRYGRRTKLYEAHSDVRESLARFIPPTIGPVQATTCCELYELVEAVENKGGDSAYL 1371	
DB	1307	STLVGRYGRRTKLYEAHSDVRESLARFIPPTIGPVQATTCCELYELVEAVENKGGDSAYL 1366	
QY	1372	EILDLCROVSRITTEPOKXONKFTTGTTIAGKVGQGISAMSKFPCALFGPMFRAIEKEIL 1431	
DB	1367	EILDLCROVSRITTEPOKXONKFTTGTTIAGKVGQGISAMSKFPCALFGPMFRAIEKEIL 1426	
QY	1432	ALLPNIPEFGDAVEESVFAAASGAGSCAVFENDEFSEPDSTONNFSLGLECVMEBCGNP 1491	
DB	1427	ALLPNIPEFGDAVEESVFAAASGAGSCAVFENDEFSEPDSTONNFSLGLECVMEBCGNP 1486	
QY	1492	QMLIRLYHLVRSAMILQAPRESLKGFWKHSGBEPTLLMNTVYNNMAIIAHCEEFDFRYA 1551	
DB	1487	QMLIRLYHLVRSAMILQAPRESLKGFWKHSGBEPTLLMNTVYNNMAIIAHCEEFDFRYA 1546	
QY	1552	ARKGDSYVLGCDYRGRMAAALLACGCKLTVDPYRITLYGVVYAPRLGTLPRDVRRA 1611	
DB	1547	ARKGDSYVLGCDYRGRMAAALLACGCKLTVDPYRITLYGVVYAPRLGTLPRDVRRA 1606	
QY	1612	GRLEKNNMGPERABEOLALACDFIRGLTNTVAOYCVDVAVSRVYSGPSGLVHNLIGMLQT 1671	
DB	1607	GRLEKNNMGPERABEOLALACDFIRGLTNTVAOYCVDVAVSRVYSGPSGLVHNLIGMLQT 1666	
QY	1672	IADGKAHFTETIKPVLDTNLSITIQVE 1698	
DB	1667	IADGKAHFTETIKPVLDTNLSITIQVE 1693	
RESULT	11		
Q9WC28			
ID	Q9WC28	PRELIMINARY; PRT; 1693 AA.	
AC	Q9WC28;		
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
RT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
RL	Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF076239; AAC27934.2; -		
DR	InterPro: IPR000606; Viral_helicase1.		
DR	InterPro: IPR002588; V_methyltransf.		
DR	InterPro: IPR002589; DUF27.		
DR	Pfam: PF01443; Viral_helicase1; 1.		
DR	Pfam: PF01660; Vmethyltransf; 1.		
DR	Pfam: PF01661; DUF27; 1.		
DR	SMART; SM00506; Alpp; 1.		
SO	SEQUENCE 1693 AA; 185578 MW; FB37B87BF6BBB501 CRC64;		
Query Match	81.4%: Score 7333; DB 12; Length 1693;		
Best Local Similarity	79.8%: Pred. No. 0;		
Matches 1375: Conservative 120; Mismatches 163; Indels 66; Gaps			


```
Db 70 VINHELELCARSGRCLEIGAHPRSLINDPNVHRCFLRPVGRDQVWYTAPTRGPAAN 129
Qy 121 CRSSALRGLEPPADRTYCFDGFSCAPAEETGVALYSLHDIIMPDAVEAMARHGXTRLYAA 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 CRSSALRGLEPPADRTYCFDGFSCAPAEETGVALYSLHDIIMPDAVEAMARHGXTRLYAA 189
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 181 LHLPPVLLPQGYHTTSVLLIHDGRVAVTYEGDTSAGCNHDSVLTIRARTKTYGDH 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 LHLPPVLLPQGYHTTSVLLIHDGRVAVTYEGDTSAGCNHDSVLTIRARTKTYGDH 249
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 241 PLVIEVRAGHGFVLLTAAPESPMPYVPYPRSTEVYRSIFGPGSGSLTPSACSTK 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 250 PLVIEVRAGHGFVLLTAAPESPMPYVPYPRSTEVYRSIFGPGSGSLTPSACSTK 309
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 301 STFHAVPVHIMDRMLFGATLDQAFCSSRLMTYLRGISYKVTGALVANEGNASDAL 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 STFHAVPVHIMDRMLFGATLDQAFCSSRLMTYLRGISYKVTGALVANEGNASDAL 369
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 361 TAITTAAYLTICHORYLRTQALISGMRLGVEHAOKFITPLYSWLPKSGRDYTPGRQL 420
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 370 TAITTAAYLTICHORYLRTQALISGMRLGVEHAOKFITPLYSWLPKSGRDYTPGRQL 429
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 421 FYACRRMLSAGHLDPRVLPFDESVPQRCRTFLKVAAGFCGFMRLGQECFCELPAP 480
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 430 FYACRRMLSAGHLDPRVLPFDESVPQRCRTFLKVAAGFCGFMRLGQECFCELPAP 489
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 481 GLVDHGHNEAVEGSEVDAEPAHLDVSGTYAVHGHOLEALYRALNPODILARASRLT 540
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 490 GLVDHGHNEAVEGSEVDAEPAHLDVSGTYAVHGHOLEALYRALNPODILARASRLT 549
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 541 ATVEIYASPRLECRVYLGKTKRTTYVVDNAHLEANGPEBYVLSFDASROSMAGSLSLT 600
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 550 ATVEIYASPRLECRVYLGKTKRTTYVVDNAHLEANGPEBYVLSFDASROSMAGSLSLT 609
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 601 YELTPAGLOYRISNGIDCATFPFGGAPSAAPCEVAAPFASALYRYARFPORSLTGLW 660
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 610 YELTPAGLOYRISNGIDCATFPFGGAPSAAPCEVAAPFASALYRYARFPORSLTGLW 669
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 661 LHPGGLGTFPPSPGHIMESANPCGEGTLTYRTWS-----TSGFSSDFSPPE 709
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 670 FHPGGLGTFPPSPGHIMESANPCGEGTLTYRTWS-----TSGFSSDFSPPE 726
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 710 AAMPAAATPGLPHSTPPVSDIVLPPSEFOVDAAVPVPAADPAGLPQPVVLTLPPPP 769
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 727 --TPSRAPPTL-----AAPLPP-----LAPDPS 749
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 770 PVHKPSIAPP-----SRRRLTYTPDGAAVVAAGSLFESDCDMLVNASNP 814
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 750 PSSAPALDEPASATSGVPAITHTOTARHRLFTYPDGSGVFAGSLFESTCTMLVNASNV 809
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 815 GHRPGLGCHAFYQRPPEAFPTPEFIMREGIAAYTLTPRILIAHAVADYVEEONPKRLEA 874
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 810 DHPGGLGCHAFYQRPPEAFPTPEFIMREGIAAYTLTPRILIAHAVADYVEEONPKRLEA 869
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 875 AYETCSRKRTAAYPLIGSITIOYVPSLSDANERNRHPDELYLTPAANWEPAKPAQ 934
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 870 AYETCSRKRTAAYPLIGSITIOYVPSLSDANERNRHPDELYLTPAANWEPAKPAQ 929
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 935 PVLTIEDTARTANLAEIDAATEVGRACAGTISPGIVAYOFTAGVPGSGKSSTIOGD 994
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 930 PVLTIEDTARTANLAEIDAATEVGRACAGTISPGIVAYOFTAGVPGSGKSSTIOGD 989
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 995 VDVVVVPTRELNRSWRRRGFAFTPHTAARVTIGRRVVIDEAPSLPHLLLHMQRASV 1054
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 990 VDVVVVPTRELNRSWRRRGFAFTPHTAARVTIGRRVVIDEAPSLPHLLLHMQRASV 1049
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1055 HLLGDPQIDPAIDFENAGLVPAIRPELAPTSWAXVTHRCADYVCELLRGAYPKIQTTSRV 1114
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1050 HLLGDPQIDPAIDFENAGLVPAIRPELAPTSWAXVTHRCADYVCELLRGAYPKIQTTSRV 1109
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1115 LRSLEWNEPAIGOKLVYTOAKANPGCAITVHEOGATFETETIITADARGLIOSSRAH 1174
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1110 LRSLEWNEPAIGOKLVYTOAKANPGCAITVHEOGATFETETIITADARGLIOSSRAH 1169
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Qy 1175 AYAALRHTREKCYITLDAPGLLREVGISDYVNNFFLAGGEVGHRSVPVIRGNPDONLGT 1234
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1170 AYAALRHTREKCYITLDAPGLLREVGISDYVNNFFLAGGEVGHRSVPVIRGNPDONLGT 1229
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1235 LOAFPSCQISAYHOLAEEELGHRPAPAAVLPCEPELEOGLVMPOLVYSDSLVELT 1294
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1230 LOAFPSCQISAYHOLAEEELGHRPAPAAVLPCEPELEOGLVMPOLVYSDSLVELT 1289
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1295 DIVCHMAAPSORKAVLSTLVGRGRRTKLYEAHSDVRESLARFTPTIGPVATTCELY 1354
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1290 DIVCHMAAPSORKAVLSTLVGRGRRTKLYEAHSDVRESLARFTPTIGPVATTCELY 1349
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1355 ELVEAMVEKQODSAYLELDLCNRDYSRTFPOKXCKFTTGETIAGKVGQGISAMSKT 1414
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1350 ELVEAMVEKQODSAYLELDLCNRDYSRTFPOKXCKFTTGETIAGKVGQGISAMSKT 1409
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1415 FCALFGFWFAIKEITALLPNIIFYGDAAVEESVFAAAGSAGSCWFEFDSFSTON 1474
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1410 FCALFGFWFAIKEITALLPNIIFYGDAAVEESVFAAAGSAGSCWFEFDSFSTON 1469
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1475 NFSGLGECVVMEECGMPQWLIRLYHLVRSAMIIQAPKESILKGFWMKSHGEPGLMNTYV 1534
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1470 NFSGLGECVVMEECGMPQWLIRLYHLVRSAMIIQAPKESILKGFWMKSHGEPGLMNTYV 1529
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1535 NMAIIACHVEFDEPVAAPFAGGDSVVLCSQYROSNNAAALIIAGCGLKLKVDYRPIGLYAG 1594
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1530 NMAIIACHVEFDEPVAAPFAGGDSVVLCSQYROSNNAAALIIAGCGLKLKVDYRPIGLYAG 1589
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1595 VVAAPGIGTLPDVYRFPARISEKNMGPGPERAROLRAVODPFRGLTNVNOVCVDVYSRV 1654
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1590 VVAAPGIGTLPDVYRFPARISEKNMGPGPERAROLRAVODPFRGLTNVNOVCVDVYSRV 1649
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1655 YGVSPGLVHNLIGMLQTIADGKAFFETIRKPVLDITNSIIQRYE 1698
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1650 YGVSPGLVHNLIGMLQTIADGKAFFETIRKPVLDITNSIIQRYE 1693
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
RESULT 12
039221
ID 039221 PRELIMINARY; PRT: 1693 AA.
AC
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
ON NCBI_Taxid=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HYDERABAD, INDIA;
RA Ansari I.H., Nanda S.K., Durgapal H., Jameel S., Panda S.K.;
RT "Eukaryotic expression of nonstructural protein (ORF1) of HEV: absence
RT of any processing.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028091; AAB82002.2; -
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; Viral_helicase1.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
DR Pfam: PF01661; DUF27; 1.
DR SMART: SM00506; Altp; 1.
KW Polyprotein.
SQ
SEQUENCE 1693 AA: 185562 MW: 7AFED003532E12BC3 CRC64:
```

```
Query Match 81.3%; Score 7329; DB 12; Length 1693;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 1374; Conservative 122; Mismatches 163; Indels 66; Gaps 6;
Qy 1 PGITTAIEQAAALAAANSALANAVVVRPFLSRVQTEIILINMOPROLVFRPEVLMNHPIOR 60
```

```

10 PGITVEGAALATANSALANAVVRPLSHQIIEILLINMOPROLVFREVEWMNPOR 69
61 VINELEOXCRABARCILEVGAHPRISINDPNVLRHCFILRPVGRDVORWYSAPTRPAAN 120
70 VINELLELCRAASRGCLIEGAHPRISINDPNVLRHCFILRPVGRDVORWYSAPTRPAAN 129
121 CRRSALRGPLPADRYCFCDFGSCAPAEATGVALYSLHDLMPADVAEAMARHGXTLYAA 180
130 CRRSALRGPLPADRYCFCDFGSCAPAEATGVALYSLHDLMPADVAEAMARHGXTLYAA 189
181 LHPPEVLLPBGTYHTTSLIHDGRVAVTYEGDTAGNHDVSLMRSTRTKVTGCH 240
190 LHPPEVLLPBGTYHTTSLIHDGRVAVTYEGDTAGNHDVSLMRSTRTKVTGCH 249
241 PLVIEVRALICGHFVLLTLTAAPESPMPYVPYRSTEVYVRSIFGSGSPLEPASCSTK 300
250 PLVIEVRALICGHFVLLTLTAAPESPMPYVPYRSTEVYVRSIFGSGSPLEPASCSTK 309
301 STFHAVPVHIMDRMLFGATLDDQAFCCSRMLTYLNGISYKVTGVALVANEGNASDAL 360
310 STFHAVPVHIMDRMLFGATLDDQAFCCSRMLTYLNGISYKVTGVALVANEGNASDAL 369
361 TXXITFAVLTICORVLRTOAISKGBRLGVEHAKETRLYSMLFEKSGROYIPGRLO 420
370 TAVITTAAYLTICORVLRTOAISKGBRLGVEHAKETRLYSMLFEKSGROYIPGRLO 429
421 FYAACRRMLSAGFHLDPRLVFEDESVPCCRTFLKRVAGKFCFEMRLGOECTCFLEPAE 480
430 FYAACRRMLSAGFHLDPRLVFEDESVPCCRTFLKRVAGKFCFEMRLGOECTCFLEPAE 489
481 GLVGDHNDNEAVEGSEVDPAPBAHLVSGTAVHGHQLEBALYRALNPODIARAARSLT 540
490 GYVGDGHNDNEAVEGSEVDPAPBAHLVSGTAVHGHQLEBALYRALNPODIARAARSLT 549
541 AATVEIYASPDLECRVTLCNKTRFTVYDGAHLEANGPROVYLSFSPASQSMGASHLT 600
550 AATVEIYASPDLECRVTLCNKTRFTVYDGAHLEANGPROVYLSFSPASQSMGASHLT 609
601 YELTPAGLOVRISNGLDCTATFPFGAPSAAPGEVAAFCSALYRNRTQHRSLTGLM 660
610 YASAGLEVRVYGAOLDHRAIFAPGVSPRNSNGEYATACSAALYRNREARHSLGNLM 669
661 LHPPEGLIGTFPPSPGHIMESANPFCGEGTLTYRTWS-----TSFSSDSFPE 709
670 FHPPEGLIGTFPPSPGHIMESANPFCGEGTLTYRTWS-----TSFSSDSFPE 726
710 AABAPAAATPGLPHSTPPVSDIMVLRPPESEFOVDAAPRPAPDAGLPGPVLTTPPPP 769
727 --TPSRAATPVL-----AALPVP-----LADPSP 749
770 PVHKPSIPPP-----SRNRLLTYTPDGAKEYAGSLFESDCMLVNASNP 814
750 PSSAPALDEBPASATSGVPAITHOTARHRLFTYDGSKVRAGSLFESTCMLVNASNP 809
815 GHPGGGLCHAFYOREPEAFYPTETIEMREGLAAYTLTPRPIIHAAVDYRVBONPKRLEA 874
810 DHPGGGLCHAFYOREPEAFYPTETIEMREGLAAYTLTPRPIIHAAVDYRVBONPKRLEA 869
875 AATRECSRGSTAYPPLGSGIYOVPRYSLFDAEMERHNRGDELYLTERPAANFEAKPKAO 934
870 AATRECSRGSTAYPPLGSGIYOVPRYSLFDAEMERHNRGDELYLTERPAANFEAKPKAO 929
935 PLUTTEDPARTANLALIEDATVEGRACAGCTISPGIYHOFYFTAGVPSSGSRSTIOOD 994
930 PLUTTEDPARTANLALIEDATVEGRACAGCTISPGIYHOFYFTAGVPSSGSRSTIOOD 989
995 VDVVVVPTRELNRSMRRGFAFTPHTAARVITGRRVVIDEAPSLPHLLLHMQRASSY 1054
990 VDVVVVPTRELNRSMRRGFAFTPHTAARVITGRRVVIDEAPSLPHLLLHMQRASSY 1049
1055 HLLGDPNOIPALIDFEPAGLVPAIRBELAPTSMWXYTHRCPADVCELIRGAYKIOQTSRY 1114

```

```

1050 HLLGDPNOIPALIDFEPAGLVPAIRBELAPTSMWXYTHRCPADVCELIRGAYKIOQTSRY 1109
1115 LSLFENPEALIGOKLVXTOAAKAMPATVHEAGATFETETIATADARGLIOSSRAH 1174
1110 LSLFENPEALIGOKLVXTOAAKAMPATVHEAGATFETETIATADARGLIOSSRAH 1169
1175 AVALTRHTEKCVIILAPGLREVGISDVIYNNFVLAGEVGNHRSVIRGNDONLCT 1234
1170 AVALTRHTEKCVIILAPGLREVGISDVIYNNFVLAGEVGNHRSVIRGNDONLCT 1229
1235 LQAFPPSCOISAVHOLAEELGHRPAPVAANLPPCPLEBGLLYMPELIVSSVLYEELT 1294
1230 LQAFPPSCOISAVHOLAEELGHRPAPVAANLPPCPLEBGLLYMPELIVSSVLYEELT 1289
1295 DIVHCRMAAPSORKAVLSTLVGRYGRRTLYEAHSDVESLARFPTIGPVQATTCELY 1354
1290 DIVHCRMAAPSORKAVLSTLVGRYGRRTLYEAHSDVESLARFPTIGPVQATTCELY 1349
1355 ELVEAMVERGODGSAVLEIDLGNRDVSRITFFOKXCNKRTTGTETAHGKVGOGISAMSKT 1414
1350 ELVEAMVERGODGSAVLEIDLGNRDVSRITFFOKXCNKRTTGTETAHGKVGOGISAMSKT 1409
1415 FCALGEPMPERATIEKELLALPNIPYGDAYEESVFAAASGASGVFEENDESSEPSTON 1474
1410 FCALGEPMPERATIEKELLALPNIPYGDAYEESVFAAASGASGVFEENDESSEPSTON 1469
1475 NFSLGLCEVMECEGMPQWILRLYLVRSAWILQAPKESLKGFWKKHSGEPGLTMMTYW 1534
1470 NFSLGLCEVMECEGMPQWILRLYLVRSAWILQAPKESLKGFWKKHSGEPGLTMMTYW 1529
1535 NNAITACHCEYDFRVAARFGDSDVYLCSDYROSNAALINGCGIKLVDRPGLVAG 1594
1530 NNAITACHCEYDFRVAARFGDSDVYLCSDYROSNAALINGCGIKLVDRPGLVAG 1589
1595 VYVAPGLTFLPVVRPAGRLSEKNMGPERBEPOLRLAICDPLRGITTNVAYOVDPVVSAY 1654
1590 VYVAPGLTFLPVVRPAGRLSEKNMGPERBEPOLRLAICDPLRGITTNVAYOVDPVVSAY 1649
1655 YGVSPGLVNLIGMLQTTIADGRAHFTETIKPVLDTNSTIIOVE 1698
1650 YGVSPGLVNLIGMLQTTIADGRAHFTETIKPVLDTNSTIIOVE 1693

RESULT 13
081344
ID 081344 PRELIMINARY: PRT: 1693 AA.
AC 081344:
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE HEPATITIS E VIRUS COMPLETE GENOME.
OS Hepatitis E virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_Taxid=12461;
RN [1]
RN SEQUENCE FROM N. A.
RC STRAIN=HEBEI;
RX MEDLINE=93348763; PubMed=8346669;
RA Bi S.L., Purdy M.A., McCausland K.A., Margolis H.S., Bradley D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China.";
RL Virus Res. 28:233-247(1993).
DR EMBL: M94177; AAA96139.1;
DR InterPro: IPR000606; Viral_helicasel.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicasel; 1.
DR Pfam: PF01660; Methyltransf; 1.
DR SMART: SM00506; A1pp; 1.
SO SEQUENCE 1693 AA; 165271 MW; E33D8F128E0B9C6F CRC64;

```

Query Match 81.0%; Score 7304; DB 12; Length 1693;
 Best Local Similarity 80.8%; Pred. No. 0;
 Matches 1380; Conservative 103; Mismatches 192; Indels 32; Gaps 7;

```

OY 1 PGTTAIEQAALAAANSALANAVVBPFLSRVOTELLINIMOPROQLVFRPEVLNHPDIQR 60
DB 10 PGTTAIEQAALAAANSALANAVVBPFLSHQOIEILLINIMOPROQLVFRPEVFNHPDIQR 69
OY 61 VINHELOYCRANAGRCLEVGAPHRSTINDPNVNLHRCFLRPVGRDQWRYSATRGPAAN 120
DB 70 VINHELELCRARSRCLEGAHPRSTINDPNVNHRCFLRPARDQWRMTAETRGPAAN 129
OY 121 CRSSALRGPLPAPARTYCFDEFSRCAPAEETGVALYSLHDMADVAEAMRHGCTRYYAA 180
DB 130 CRSSALRGPLPADRTYCFDEFSRCNPAETGVALYSLHDMSPDVAEAMRHGCTRYYAA 189
OY 181 LHLPEVLLPPTGTYHTSTYLLIHDGDAVVTYEGDTSAGYNHDSILRAMIRTKIYGDH 240
DB 190 LHLPEVLLPPTGTYHTSTYLLIHDGDAVVTYEGDTSAGYNHDSILRAMIRTKIYGDH 249
OY 241 PLYIERVRAIGCHFVLLITAPPEPMPYVYPRSTEVYRSTJFPGGSPSLPSSACSTK 300
DB 250 PLYIERVRAIGCHFVLLITAPPEPMPYVYPRSTEVYRSTJFPGGSPSLPSSACSTK 309
OY 301 STEHAVPVHIMDRLMLGATLDDOAFCCSRLMTYLGISKYVGVALYANEGMNASDAL 360
DB 310 STEHAVPVHIMDRLMLGATLDDOAFCCSRLMTYLGISKYVGVALYANEGMNASDAL 369
OY 361 TAXITAYLITICHOYRLTQOAIKGMRLGVEHAOKFITLYSWLFESGRDIYPGHOLQ 420
DB 370 TAXITAYLITICHOYRLTQOAIKGMRLGVEHAOKFITLYSWLFESGRDIYPGHOLQ 429
OY 421 FYACRRMLTSGAHLDPRVLVDESVPCRCRTEFLKVAAGFCFPMRLGECFCELEPAE 480
DB 430 FYACRRMLTSGAHLDPRVLVDESVPCRCRTEFLKVAAGFCFPMRLGECFCELEPAE 489
OY 481 GLVGDHNDHEAYEGSVDPADPEAHLDVSGTYAVHGHOLALYALRALVPODIAARSLT 540
DB 490 GLVGDHNDHEAYEGSVDPADPEAHLDVSGTYAVHGHOLALYALRALVPODIAARSLT 549
OY 541 ATVELVASPDLRCLRVNLGNKRTFTTVVDAHLLEANGPEOYVLSFDSASROSMGASHSLT 600
DB 550 ATVELVASPDLRCLRVNLGNKRTFTTVVDAHLLEANGPEOYVLSFDSASROSMGASHSLT 609
OY 601 YELTPAGLOVRISNGIDCTATPPPGCAPSAPEVAFCALYRYNRFTQHSJTGGLW 660
DB 610 YELTPAGLOVRISNGIDCTATPPPGCAPSAPEVAFCALYRYNRFTQHSJTGGLW 669
OY 661 LHPEGLIGTPPPSPGHIMESANPFCGEGTLTYRTWS-TJGFSDFSP-----PEAAPA 714
DB 670 LHPEGLIGTPPPSPGHIMESANPFCGEGTLTYRTWS-TJGFSDFSP-----PEAAPA 729
OY 715 MAATPGLPHSTPRYSITWLVPPSEEPQVDAAPV-PPAPPAGLV-GRVVLTPPPPPVH 772
DB 730 MAATPGLPHSTPRYSITWLVPPSEEPQVDAAPV-PPAPPAGLV-GRVVLTPPPPPVH 766
OY 773 KPSTLP-PPSNRRLLTYTPGAKYVAGSLFESDCLVYANSNGCHRRGGCICIAFYRFP 831
DB 767 KPSTLP-PPSNRRLLTYTPGAKYVAGSLFESDCLVYANSNGCHRRGGCICIAFYRFP 826
OY 832 EAPYPTFEIRREGLAATYTLPRPIIHAVADYRVEONPKRLLEAAYRETCSRGAAPPL 891
DB 827 EAPYPTFEIRREGLAATYTLPRPIIHAVADYRVEONPKRLLEAAYRETCSRGAAPPL 886
OY 892 GSGIYOYVSLSPDAMERNHRPGDELITLPEANMFEANPAPQVLTITEDTARTANIAL 951
DB 887 GSGIYOYVSLSPDAMERNHRPGDELITLPEANMFEANPAPQVLTITEDTARTANIAL 946
OY 952 EIDATATVGRACAGCTISPGIVHQTAFAGVSGSKSRISIOGDVYVYVPTRELNRSMR 1011
DB 947 EIDATATVGRACAGCTISPGIVHQTAFAGVSGSKSRISIOGDVYVYVPTRELNRSMR 1006
OY 1012 RGFAAFTPHTAARVTIGRRVVIDEAPSLPHLLLHMORASSVHLIDDPNQIPAIIDEHA 1071
  
```

```

DB 1007 RGFAAFTPHTAARVTIGRRVVIDEAPSLPHLLLHMORASSVHLIDDPNQIPAIIDEHA 1066
OY 1072 GLVPAIRPELAPRTSMXAVTHRCPADYCELLRGAYPKIQTTSRYLRSIFMNEPRLIGOKLY 1131
DB 1067 GLVPAIRPELAPRTSMXAVTHRCPADYCELLRGAYPKIQTTSRYLRSIFMNEPRLIGOKLY 1126
OY 1132 TOAKAANPGATIVHEAGSTFETTTIATADARGLIOSSRAHAVALTHRTKCYTLDA 1191
DB 1127 TOAKAANPGATIVHEAGSTFETTTIATADARGLIOSSRAHAVALTHRTKCYTLDA 1186
OY 1192 PGLIREGISDVIYVNEFFLAGGEVGHKRPVIRGNPDQMLGTLOAFPPSCQISAYHOLA 1251
DB 1187 PGLIREGISDVIYVNEFFLAGGEVGHKRPVIRGNPDQMLGTLOAFPPSCQISAYHOLA 1246
OY 1252 EELGHRPAPAAVLPCCPELEGGLLYPOELTTCDSVTFELNDIVHCRMAAPSORKAVL 1311
DB 1247 EELGHRPAPAAVLPCCPELEGGLLYPOELTTCDSVTFELNDIVHCRMAAPSORKAVL 1306
OY 1312 STLVRGRTKLYEAHSDVRESIAFTPTIGPVATTCETELVEYAMVEKODGSAYL 1371
DB 1307 STLVRGRTKLYEAHSDVRESIAFTPTIGPVATTCETELVEYAMVEKODGSAYL 1366
OY 1372 ELDLCNRDVSRIITFFOKXCKFTTGETIAGKYGOGISAMSKTFPCALFGMPRAIEREIL 1431
DB 1367 ELDLCNRDVSRIITFFOKXCKFTTGETIAGKYGOGISAMSKTFPCALFGMPRAIEREIL 1426
OY 1432 ALLPPIIFYGDAYEESYFAAASGAGSCWFEENDESEFSTONNFSLGLEBCVMECGMP 1491
DB 1427 ALLPPIIFYGDAYEESYFAAASGAGSCWFEENDESEFSTONNFSLGLEBCVMECGMP 1486
OY 1492 QMLIRLHLVRSAMIIQAPRESLKGFWKHSRGEGTLLMNTVMNMAIIACIYERDPRVA 1551
DB 1487 QMLIRLHLVRSAMIIQAPRESLKGFWKHSRGEGTLLMNTVMNMAIIACIYERDPRVA 1546
OY 1552 AFKGDSDVLCSDVROSANAAIIAGCGLKLYDYRIGIYAGVVAAPGLGTLPDVVRFA 1611
DB 1547 AFKGDSDVLCSDVROSANAAIIAGCGLKLYDYRIGIYAGVVAAPGLGTLPDVVRFA 1606
OY 1612 GRLEKKNWGPGERAEQRLAVCDFTLGLTNVAVCVVVSRYGVSPGLVHNLIGMLQT 1671
DB 1607 GRLEKKNWGPGERAEQRLAVCDFTLGLTNVAVCVVVSRYGVSPGLVHNLIGMLQT 1666
OY 1672 IADGKAFTETIKPVLDLTNSITQRE 1698
DB 1667 IADGKAFTETIKPVLDLTNSITQRE 1693

```

RESULT 14
 069418
 ID 069418; PRELIMINARY; PRT; 1693 AA.
 AC 069418;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ORF1, ORF2 & ORF3.
 OS Hepatitis E virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.
 OX NCBI_Taxid=12461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=INDIVIDUAL PATIENT INFECTED WITH HEV;
 RA Von Brunn A., Seebach J., Thyagarajan S.P., Mohanavalli B., Menon T.,
 Froesner G.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X99441; CAA67802.1; -
 DR InterPro: IPR000606; Viral helicase1.
 DR InterPro: IPR002588; V_methyltransf.
 DR InterPro: IPR002589; DUF27.
 DR Pfam: PF01443; Viral_helicase1, 1.
 DR Pfam: PF01660; Vmethyltransf; 1.
 DR Pfam: PF01661; DUF27; 1.
 DR SMART: SM00506; Alpp; 1.

SQ SEQUENCE 1693 AA: 185949 MM: 489B4ADBA5E7E529 CRC64:
 Query Match 81.0%; Score 7301; DB 12; Length 1693;
 Best Local Similarity 79.9%; Pred. No. 0;
 Matches 1367; Conservative 116; Mismatches 188; Indels 40; Gaps 6;

QY 1 PGTTTAEQAALAAASALANAVVRPLSRVOTELLINMQPRLQVREPLVNHPIOR 60
 DB PGTTTTAEQAALAAASALANAVVRPLSRVOTELLINMQPRLQVREPLVNHPIOR 69
 QY 61 VHNLELCYCRARAGCIEGAGHRSINPNVNLHRCFLRPVGRDQVRYMSATRRPAN 120
 DB VHNLELCYCRARAGCIEGAGHRSINPNVNLHRCFLRPVGRDQVRYMSATRRPAN 129
 QY 121 CRRSALRGPLPADRTYCFDGFSCAFAETGVALYSLHDLMPADVAEAMARHGXTLYAA 180
 DB CRRSALRGPLPADRTYCFDGFSCAFAETGVALYSLHDLMPADVAEAMARHGXTLYAA 189
 QY 130 CRRSALRGPLPADRTYCFDGFSCAFAETGVALYSLHDLMPADVAEAMARHGXTLYAA 189
 DB CRRSALRGPLPADRTYCFDGFSCAFAETGVALYSLHDLMPADVAEAMARHGXTLYAA 189
 QY 181 LHLPEVLLPFGTYHTTSTYLLHDDGRAVYVEGDTSGAGYNDVSTILRAMIRTKTIGDH 240
 DB LHLPEVLLPFGTYHTTSTYLLHDDGRAVYVEGDTSGAGYNDVSTILRAMIRTKTIGDH 249
 QY 190 LHLPEVLLPFGTYHTTSTYLLHDDGRAVYVEGDTSGAGYNDVSTILRAMIRTKTIGDH 249
 DB LHLPEVLLPFGTYHTTSTYLLHDDGRAVYVEGDTSGAGYNDVSTILRAMIRTKTIGDH 249
 QY 241 PLVTEVRATIGCFEVLTLAAAPSPMPYVPRSTEVYVRSITFGGSGSPSLFSPSCSTK 300
 DB PLVTEVRATIGCFEVLTLAAAPSPMPYVPRSTEVYVRSITFGGSGSPSLFSPSCSTK 309
 QY 250 PLVTEVRATIGCFEVLTLAAAPSPMPYVPRSTEVYVRSITFGGSGSPSLFSPSCSTK 309
 DB PLVTEVRATIGCFEVLTLAAAPSPMPYVPRSTEVYVRSITFGGSGSPSLFSPSCSTK 309
 QY 301 STEFAVPAHIMDLMEGATLDDQAFCCSRMLTYLRGISYKTVGALVANEENASSEDAL 360
 DB STEFAVPAHIMDLMEGATLDDQAFCCSRMLTYLRGISYKTVGALVANEENASSEDAL 369
 QY 310 STEFAVPAHIMDLMEGATLDDQAFCCSRMLTYLRGISYKTVGALVANEENASSEDAL 369
 DB STEFAVPAHIMDLMEGATLDDQAFCCSRMLTYLRGISYKTVGALVANEENASSEDAL 369
 QY 361 TAITTAAYLTICHORYLRTQALSKGRRLVEHAQKFTRLXSWLFEKSGRDYIPGRDQ 420
 DB TAITTAAYLTICHORYLRTQALSKGRRLVEHAQKFTRLXSWLFEKSGRDYIPGRDQ 429
 QY 370 TAITTAAYLTICHORYLRTQALSKGRRLVEHAQKFTRLXSWLFEKSGRDYIPGRDQ 429
 DB TAITTAAYLTICHORYLRTQALSKGRRLVEHAQKFTRLXSWLFEKSGRDYIPGRDQ 429
 QY 421 FTAQCRRLISAGHLDPRVLYNDESPCCRTFLKRYAKVAFCCFMWLOECGCFLEPAE 480
 DB FTAQCRRLISAGHLDPRVLYNDESPCCRTFLKRYAKVAFCCFMWLOECGCFLEPAE 489
 QY 430 FTAQCRRLISAGHLDPRVLYNDESPCCRTFLKRYAKVAFCCFMWLOECGCFLEPAE 489
 DB FTAQCRRLISAGHLDPRVLYNDESPCCRTFLKRYAKVAFCCFMWLOECGCFLEPAE 489
 QY 481 GLVGDHGHNEAYEGSEVDPAEPANLDSGYAVHGHOLEALYBALNVODIJAARSRIT 540
 DB GLVGDHGHNEAYEGSEVDPAEPANLDSGYAVHGHOLEALYBALNVODIJAARSRIT 549
 QY 490 GLVGDHGHNEAYEGSEVDPAEPANLDSGYAVHGHOLEALYBALNVODIJAARSRIT 549
 DB GLVGDHGHNEAYEGSEVDPAEPANLDSGYAVHGHOLEALYBALNVODIJAARSRIT 549
 QY 541 AVEELVAPDRLECRITVGNKTFRTTVGAMLEANGPEOYVLSFSDASROSMAGSHSLT 600
 DB AVEELVAPDRLECRITVGNKTFRTTVGAMLEANGPEOYVLSFSDASROSMAGSHSLT 609
 QY 550 AVEELVAPDRLECRITVGNKTFRTTVGAMLEANGPEOYVLSFSDASROSMAGSHSLT 609
 DB AVEELVAPDRLECRITVGNKTFRTTVGAMLEANGPEOYVLSFSDASROSMAGSHSLT 609
 QY 601 YELTPAGLOVRISNGLDCTATPPGAPASAPAGEVAACALYRYNRTORSHSLTGIM 660
 DB YELTPAGLOVRISNGLDCTATPPGAPASAPAGEVAACALYRYNRTORSHSLTGIM 669
 QY 610 YAVSADGLEVRVAAGLDRAVAFAPGVSPTGEGTATACALYRRENVORSHSLTGIM 669
 DB YAVSADGLEVRVAAGLDRAVAFAPGVSPTGEGTATACALYRRENVORSHSLTGIM 669
 QY 661 LHPGGLIGFPFPGHIMESANPFCGEGTLTTRWS--TSGFSSDRSP-----PEAAAPA 714
 DB LHPGGLIGFPFPGHIMESANPFCGEGTLTTRWS--TSGFSSDRSP-----PEAAAPA 729
 QY 670 LHPGGLIGFPFPGHIMESANPFCGEGTLTTRWS--TSGFSSDRSP-----PEAAAPA 729
 DB LHPGGLIGFPFPGHIMESANPFCGEGTLTTRWS--TSGFSSDRSP-----PEAAAPA 729
 QY 715 MATPGCLPHSTPVSIDWVLPPESEFQYDAV--PPAPDPAGLPGVVLTPPPPPVHK 773
 DB MATPGCLPHSTPVSIDWVLPPESEFQYDAV--PPAPDPAGLPGVVLTPPPPPVHK 762
 QY 730 RAATPFL-----AAFLPPAPDPAPPPS-----AAPADPPASG 762
 DB RAATPFL-----AAFLPPAPDPAPPPS-----AAPADPPASG 762
 QY 774 PSITPP-----SRNRLLYTPDCAKAVAGSLFESDCMLVYNASPGHRPGGGLCHAFY 827
 DB PSITPP-----SRNRLLYTPDCAKAVAGSLFESDCMLVYNASPGHRPGGGLCHAFY 822
 QY 763 TTAGAPAITHOMARRRLFTYPDGSKYFAGSLFESTCTWPNVNASVNDHRPGGGLCHAFY 822
 DB TTAGAPAITHOMARRRLFTYPDGSKYFAGSLFESTCTWPNVNASVNDHRPGGGLCHAFY 822
 QY 828 GREPEAFYTERFIMREGLAAYTLTPRLTHAVALDYRVONKRLLEAAREFCRSRGTAA 887
 DB GREPEAFYTERFIMREGLAAYTLTPRLTHAVALDYRVONKRLLEAAREFCRSRGTAA 882
 QY 832 QRYPSAFDAFVNRGAAYTLTPRLTHAVALDYRVONKRLLEAAREFCRSRGTAA 882
 DB QRYPSAFDAFVNRGAAYTLTPRLTHAVALDYRVONKRLLEAAREFCRSRGTAA 882
 QY 888 YPLLSGTYQVPSLSFDMERNHRRGDELVLEPAANFEANKPAQVLTITTEDTARTAA 947
 DB YPLLSGTYQVPSLSFDMERNHRRGDELVLEPAANFEANKPAQVLTITTEDTARTAA 942
 QY 883 YPLLSGTYQVPSLSFDMERNHRRGDELVLEPAANFEANKPAQVLTITTEDTARTAA 942
 DB YPLLSGTYQVPSLSFDMERNHRRGDELVLEPAANFEANKPAQVLTITTEDTARTAA 942
 QY 948 NLALEIDATENVACAGCTISPGIVHVOFTAGVPGSGKSRISIQGGDVVVVPTRELN 1007
 DB NLALEIDATENVACAGCTISPGIVHVOFTAGVPGSGKSRISIQGGDVVVVPTRELN 1007

DB 943 NLALEIDATENVACAGCTISPGIVHVOFTAGVPGSGKSRISIQGGDVVVVPTRELN 1002
 QY 1008 SWRRRGFAAFTPTTAARVTTIGRRVVIDEAPSLPHLLHMQASSVHLGDPNQIPAD 1067
 DB 1003 AMRRRGFAAFTPTTAARVTTIGRRVVIDEAPSLPHLLHMQASSVHLGDPNQIPAD 1062
 QY 1068 FEHAGLVPAIRBELPTSMWXYTHRCPADVCELIRGAYKIOITTSRYLSLFWNEPAIQ 1127
 DB 1063 FEHAGLVPAIRBELPTSMWXYTHRCPADVCELIRGAYKIOITTSRYLSLFWNEPAIQ 1122
 QY 1128 KLVYQAAKAAAPGATVHEAGAFETETTTATADARGLIOSRAHVALTTRHREKV 1187
 DB 1123 KLVYQAAKAAAPGATVHEAGAFETETTTATADARGLIOSRAHVALTTRHREKV 1182
 QY 1188 ILDAPGLREVGISDITVNNPFLAGGVXGHHPSVYPRGNPNQNLGTLQAPFSCQISAY 1247
 DB 1183 ILDAPGLREVGISDITVNNPFLAGGVXGHHPSVYPRGNPNQNLGTLQAPFSCQISAY 1242
 QY 1248 HOLAEELHRRPAPVAALPCEPELEOGLLYMPQELTVSDSVLFELTDIVHCRMAAPSOR 1307
 DB 1243 HOLAEELHRRPAPVAALPCEPELEOGLLYMPQELTVSDSVLFELTDIVHCRMAAPSOR 1302
 QY 1308 KAVLSTLVGRVGRRTKLYAAHSVDRESLARIPITGPVQATTCETELVEAMVKKGOOG 1367
 DB 1303 KAVLSTLVGRVGRRTKLYAAHSVDRESLARIPITGPVQATTCETELVEAMVKKGOOG 1362
 QY 1368 SAVLELIDLCNRPVSRITTFPOKXCNKFTGETTAHGKVGOGISAMSKTECALGPMFRATE 1427
 DB 1363 SAVLELIDLCNRPVSRITTFPOKXCNKFTGETTAHGKVGOGISAMSKTECALGPMFRATE 1422
 QY 1428 KEILALPKNIFGDAYEESVFAAASGASGCMVEENDESEFDSQNNESLGLCECVME 1487
 DB 1423 KEILALPKNIFGDAYEESVFAAASGASGCMVEENDESEFDSQNNESLGLCECVME 1482
 QY 1488 CGMPQMLRPLVLYNSAVTLQAPKESLKGFMKKHSGEPETLLMNTVMNNAITAHCTEPD 1547
 DB 1483 CGMPQMLRPLVLYNSAVTLQAPKESLKGFMKKHSGEPETLLMNTVMNNAITAHCTEPD 1542
 QY 1548 FVYAAFGGDSVVLCSDYQSRNAALTAGCCLKLVDRPGLAGVYVAGLGLTDPV 1607
 DB 1543 FVYAAFGGDSVVLCSDYQSRNAALTAGCCLKLVDRPGLAGVYVAGLGLTDPV 1602
 QY 1608 VRFAGRLSEKMGPPERAQDLAVCDLRLGLTVAOVCYDVVSRVYGVSPGLVHNLG 1667
 DB 1603 VRFAGRLSEKMGPPERAQDLAVCDLRLGLTVAOVCYDVVSRVYGVSPGLVHNLG 1662
 QY 1668 MLOTTADGKAHFTETIKPYLDJNSIQRYE 1698
 DB 1663 MLOTTADGKAHFTETIKPYLDJNSIQRYE 1693
 RESULT 15
 Q81873 PRELIMINARY: PRT: 727 AA.
 AC 081873:
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE POLYPROTEIN (ENCODING NTP-BINDING PROTEIN AND RNA-DEPENDENT RNA
 DE POLYMERASE) (FRAGMENT).
 OS Hepatitis E virus.
 OC Viruses: ssRNA positive-strand viruses, no DNA stage.
 OX NCBI_Taxid:12461.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tam A.W., Smith M.M., Kim J.P., Young L.M., Platak M., Feldman R.A.,
 RA Purdy M.A., Bradley D.W., Reyes G.R., Fry K.E.,
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
 RN [2]
 RX SEQUENCE FROM N.A.
 RX MEDLINE=90193694; PubMed=2107574;
 RA Reyes G.R., Purdy M.A., Kim J.P., Luk K.C., Young L.M., Fry K.E.,
 RA Bradley D.;

RT "Isolation of a cDNA from the virus responsible for enterically
RT transmitted non-A, non-B hepatitis.";
RL Science 247:1335-1339(1990).
RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=92271462; PubMed=1589964;

RA Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,

Platak M., Feldman R.A., Yun K.Y., Purdy M.A., Et Al.;

RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GTP binding site.";

RL Virus Genes 6:173-185(1992).

DR EMBL: M32400; AAA03206.1; "

DR InterPro: IPR000606; Viral_helicase1.

DR Pfam: PF01443; Viral_helicase1.

KM Polypeptide: RNA-directed RNA polymerase.

FT NON_TER 1

SQ SEQUENCE 727 AA; 79306 MW; 872691f6c2318fa7 CRC64;

Query Match 38.2%; Score 3444; DB 12; Length 727;

Best Local Similarity 88.7%; Pred. No. 2,4e-243;

Matches 645; Conservative 40; Mismatches 42; Indels 0; Gaps 0;

QY 972 IHHYQFTAGVPGSGSKRSISIOGDVYVYVPTRELNSWRGRGFAFPPTAARTIGRRV 1031
DB 1 VVOYQFTAGVPGSGSKRSISITQADVDVYVYVPTRELNSWRGRGFAFPPTAARTIGRRV 60
QY 1032 VIDAPSLPPLLHLLHMQRASSVHLGDPNQIPAIDEFHAGLVPAIRBELAPTSWMXVTH 1091
DB 61 VIDAPSLPPLLHLLHMQRAATVHLGDPNQIPAIDEFHAGLVPAIRDLGPTSMWHYTH 120
QY 1092 RCPADVCELTRGATPKIQTTSRVLRSLFWMNPAIGOKLYTQAAKANPGAITHEAOGA 1151
DB 121 RWPADVCELTRGATPKIQTTSRVLRSLFWMNPAIGOKLYTQAAKANPGGSVTHEAOGA 180
QY 1152 TPTFTTITATPADARGLIQSSRAHAIVALTRTEKCVIIDAPGLREVGISDVIYNNFELA 1211
DB 181 TPTFTTITATPADARGLIQSSRAHAIVALTRTEKCVIIDAPGLREVGISDVIYNNFELA 240
QY 1212 GGEVGHRRPSVYIPRGNDPQNLGLQAFPPSCQISAYHQLAEELGHRPAVPAVLPCEPEL 1271
DB 241 GGEVGHRRPSVYIPRGNDPQNLGLQAFPPSCQISAYHQLAEELGHRPAVPAVLPCEPEL 300
QY 1272 EQGLLYMPQELTVSDSVLVEFLTDIVHCMAAPSQOKAVLSTLVGRYGRRTKLYEAHSD 1331
DB 301 EQGLLYMPQELTVSDSVLVEFLTDIVHCMAAPSQOKAVLSTLVGRYGRRTKLYEAHSD 360
QY 1332 VRESIARFIPITIGPVQATTCGLYELVEAMVEKGODGSVLELDCNRDVSRTFFQKXCN 1391
DB 361 VRESIARFIPITIGPVQATTCGLYELVEAMVEKGODGSVLELDCNRDVSRTFFQKXCN 420
QY 1392 KFTTGETIANGKVGQGISAMSKTFCALFGPWFRAIEKEITLALPNIIFYGAYEESVFAA 1451
DB 421 KFTTGETIANGKVGQGISAMSKTFCALFGPWFRAIEKEITLALPQGVFYGDAPFDIVFSA 480
QY 1452 AVSAGSCMVFENDSEPDSTONNFSLGLECVNMEECGMPQWLIRLYHLVRSAMITQAPK 1511
DB 481 AVSAGSCMVFENDSEPDSTONNFSLGLECVNMEECGMPQWLIRLYHLVRSAMITQAPK 540
QY 1512 ESLKGFHKHSGEGPTLLMNTVMMALIIAHCEFRDPRVAAFKGDDSVILCSDRQSRNA 1571
DB 541 ESLKGFHKHSGEGPTLLMNTVMMALIIAHCEFRDPRVAAFKGDDSVILCSDRQSRNA 600
QY 1572 AALTAGGLKLIKVDYRPIGLAGVVAAGLGLTLPDVVRFAGRLSEKNMGPPPERAEQRL 1631
DB 601 AALTAGGLKLIKVDYRPIGLAGVVAAGLGLTLPDVVRFAGRLSEKNMGPPPERAEQRL 660
QY 1632 AVCDFLGLTNVAQVYVSVRYGVSPGLVHNLIGMLQTIADGKAHFTETIKPVLDTN 1691
DB 661 AVCDFLGLTNVAQVYVSVRYGVSPGLVHNLIGMLQTIADGKAHFTETIKPVLDTN 720
QY 1692 SIIGRVE 1698

DB 721 SIIGRVE 727

Search completed: March 7, 2002, 14:08:05
Job time: 303 sec
